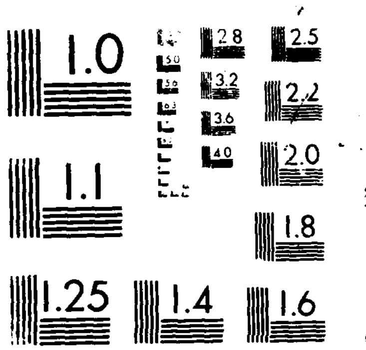
AD-A186 637 CONTROL VARIATE SELECTION FOR MULTIRESPONSE SIMULATION 1/3
(U) AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH
K W BAUER MAY 87 AFIT/CI/NR-87-132D

UNCLASSIFIED F/G 12/3 NL



RESOLUTION TEST CHA

STANDARDS 1963

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

DTIC FILE COPY

AD-A186 637

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
REPORT NUMBER AFIT/CI/NR 87-132D	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
TITLE (and Subtitle) Control Variate Selection for Multiresponse Simulation		5. TYPE OF REPORT & PERIOD COVERED THESIS/DISSERTATION
		6. PERFORMING ORG. REPORT NUMBER
AUTHOR(s) Kenneth William Bauer, Jr.		8. CONTRACT OR GRANT NUMBER(s)
PERFORMING ORGANIZATION NAME AND ADDRESS AFIT STUDENT AT: Purdue University		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS AFIT/NR WPAFB OH 45433-6583		12. REPORT DATE 1987
		13. NUMBER OF PAGES 182
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

APPROVED FOR PUBLIC RELEASE: IAW AFR 190-1

LYNN E. WOLAVER 26 Oct 87
Dean for Research and
Professional Development
AFIT/NR

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

ATTACHED

DTIC
SELECTE
NOV 2 4 1987
S D
H

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

ABSTRACT

Bauer, Kenneth W., Jr. Ph.D., Purdue University, May 1987.
Control Variate Selection for Multiresponse Simulation.
Major Professor: James R. Wilson.

A solution is offered to the general problem of optimal selection of control variates. Solutions are offered for two different cases of the general problem: ~~(a)~~ when the covariance matrix of the controls is unknown, and ~~(b)~~ when the covariance matrix of the controls is known and is incorporated into point and confidence region estimators. For the second case a new estimator is introduced. Under the assumption that the responses and the controls are jointly normal, the unbiasness of this new estimator is established , and its dispersion matrix is derived. A selection algorithm is implemented which locates the optimal subset of controls. The algorithm is based on criteria derived for the two cases listed above. A promising new class of controls is introduced which are called "routing variables". The asymptotic distribution of these controls is derived as well as their asymptotic mean and variance. Finally, the performance of the selection algorithm is investigated and the new estimator is contrasted with the classical estimator.

A-1

87 11 10 08

BIBLIOGRAPHY

Aitkin, M. A., "Simultaneous Inference and the Choice of Variable Subsets in Multiple Regression," *Technometrics*, vol. 16, pp. 221-227, 1974.

Akaike, H., "Information Theory and an Extension of the Maximum Likelihood Principle," *2nd International Symposium on Information Theory*, pp. 267-281, Akademiai Kiado, Budapest, 1973.

Allen, D. M., "Mean Square Error of Prediction as a Criterion for Selecting Variables," *Technometrics*, vol. 13, pp. 469-475, 1971.

Anderson, T. W., *An Introduction to Multivariate Statistical Analysis*, John Wiley, New York, New York, 1984.

Arvensen, J. N., "Jackknifing U-Statistics," *Annals of Mathematical Statistics*, vol. 40, pp. 2076-2100, 1969.

Bauer, Kenneth W., A Monte Carlo Study of Dimensionality Assessment and Factor Interpretation in Principal Components Analysis, Unpublished Masters Thesis, Air Force Institute of Technology, 1981.

Beaton, A. E., The Use of Special Matrix Operators in Statistical Calculus, Research Bulletin RB-64-51, Educational Testing Service, Princeton, New Jersey, 1964.

Box, George E., William G. Hunter, and J. Stuart, *Statistics for Experimenters*, John Wiley and Sons, New York, New York, 1978.

Cheng, R. C. H., "Analysis of Simulation Experiments under Normality Assumptions," *Journal of the Operational Research Society*, vol. 29, pp. 493-497, 1978.

Chvatal, Vasek, *Linear Programming*, W.H. Freeman and Company, New York, New York, 1980.

Crane, M. and A. J. Lemoine, "An Introduction to the Regenerative Method for Simulation Analysis," in *Lecture Notes in Control and Information Sciences*, Springer-Verlag, Berlin, Germany, 1977.

Draper, N. and H. Smith, *Applied Regression Analysis, Second Edition*, John Wiley and Sons, New York, New York, 1981.

Eakle, J. D., Regenerative Analysis Using Internal Controls, Unpublished Ph.D. Dissertation, Mechanical Engineering Department, University of Texas, Austin, Texas, 1982.

Flury, Bernard and Hans Reidwyl, "T ** 2 Tests and the Linear Two-Group Discriminant Function, and their Computation by Linear Regression," *The American Statistician*, vol. 39, pp. 20-25, February, 1985.

Furnival, G. M. and R. W. Wilson, "Regression by Leaps and Bounds," *Technometrics*, no. 16, pp. 499-511, 1974.

Furnival, George M., "All Possible Regressions with Less Computation," *Technometrics*, vol. 13, pp. 403-408, May 1971.

Gabriel, K. R., "Simultaneous Test Procedures - Some Theory of Multiple Comparisons," *Annals of Mathematical Statistics*, vol. 40, pp. 224-250, 1969.

Hocking, R. R., "The Analysis and Selection of Variables in Linear Regression," *Biometrics*, vol. 32, pp. 1-49, March, 1976.

Hocking, R. R., "Developments in Linear Regression Methodology: 1959-1982," *Technometrics*, vol. 25, pp. 219-230, August, 1983.

Hoerl, A. E. and R. W. Kennard, "Ridge Regression: Biased estimation for non-orthogonal problems," *Technometrics*, vol. 12, pp. 55-68, 1970.

Hogg, R. V. and A. T. Craig, *Introduction to Mathematical Statistics,* Macmillan Company, London, England, 1970.

Iglehart, D. L., "The Rengerative Method for Simulation Analysis," in *Current Trends in Programming Methodology, Vol III* , ed. K. M. Chandy and R. Yeh , Prentice Hall, Englewood Cliffs, New Jersey, 1978.

Iglehart, D. L. and P. A. W. Lewis, "Regenerative Simulation with Internal Controls," *Journal of the Association of Computing Machinery*, vol. 26, no. 2, pp. 271-282, 1979.

Johnson, R. A. and D. W. Wichern, *Applied Multivariate Analysis,* Prentice Hall, Inc. , Englewood Cliffs, New Jersey, 1982.

Jollife, I. T., "Disgarding Variables in Principal Components Analysis: Part I , Artifical Data," *Applied Statistics*, vol. 21, pp. 160-173, 1972.

Kennedy, W. J. and J. E. Gentle, *Statistical Computing,* Marcel Dekker, Inc., New York, New York, 1980.

Kenney, J. F. and E. S. Keeping, *Mathematics of Statistics Part II,* Van Nostrand Co. Inc., New York, New York, 1951.

Kleijnen, J. P. C., *Statistical Techniques in Simulation, Part 1 and 2,* Marcel Deckker, New York, New York, 1975.

Lavenberg, S. S., T. L. Moeller, and P. D. Welch, Statistical Results on Multiple Control Variables with Application to Variance Reduction in Queueing Network Simulation, IBM Research Report RC-7423, Yorktown Heights, New York, 1978.

Lavenberg, S. S., T. L. Moeller, and C. H. Sauer, "Concomitant Control Variables Applied to the Regenerative Simulation of Queueing Systems," *Operations Research*, vol. 27, no. 1, pp. 134-160, 1979.

Lavenberg, S. S. and P. D. Welch, "A Perpective on the Use of Control Variables to Increase the Efficiency of Monte Carlo Simulations," *Management Sciences*, vol. 27, pp. 322-334, March 1981.

Lavenberg, S. S., T. L. Moeller, and P. D. Welch, "Statistical Results on Control Variables with Application to Queueing Network Simulation," *Operations Research*, vol. 30, pp. 182-202, Jan-Feb 1982.

Lavenberg, Stephen S., T. L. Moeller, and P. D. Welch, "Statistical Results on Control Variables with Applications to Queueing Network Simulation," *Operation Research*, vol. 30, pp. 182-202.

Lindley, D. V., "The Choice of Variables in Multiple Regression," *Journal of the Royal Statistical Society*, vol. B30, pp. 31-53, 1968.

Mallows, C. L., "Some Comments on C sub p," *Technometrics*, vol. 15, pp. 661-675, 1973.

Mantel, N., "Why Stepdown Procedures in Variable Selection," *Technometrics*, vol. 12, pp. 621-625, 1970.

McCabe, George P., "Evaluation of Regression Coefficient Estimates Using alpha-acceptabilty," *Technometrics*, vol. 20, pp. 131-139, May 1978.

McCabe, George P., "Principal Variables," *Technometrics*, vol. 26, pp. 137-144, May, 1984.

McKay, R. J., "Variable Selection in Multivariate Regression: An Application of Simultaneous Test Procedures," *Journal of the Royal Statistical Society*, vol. B 39, pp. 371-380, 1977.

Miller, Rupert G., "The jackknife - a review," *Biometrika*, vol. 61, pp. 1-15, 1974.

Muirhead, R. J., *Aspects of Multivariate Statistical Theory,* John Wiley and Sons, New York, New York, 1982.

Neter, J., W. Wasserman, and M. H. Knuter, *Applied Linear Regression Models,* Richard D. Erwin, Inc., Homewood, Illinois, 1983.

Neuts, Marcel F., *Probability,* Allyn and Bacon, Inc., Boston, Mass., 1973.

Nova, A. M. Porta, A Generalized Approach to Variance Reduction in Discrete-event Simulation using Control Variables, Unpublished Ph.D. Dissertation, Department of Mechanical Engineering, The University of Texas, Austin, Texas, 1985.

Nozari, A., S. F. Arnold, and C. D. Pegden, "Control Variates for Multipopulation Experiments," *IIE Transactions*, vol. 16, pp. 159-169, June, 1984.

Pritsker, A. Alan B., *Introduction to Simulation and SLAM,* Halsted Press, New York, New York, 1986.

Rao, C. R., "Least Squares Theory using an Estimated Dispersion Matrix and its Application to Measurement of Signals," *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probabilty*, vol. I, pp. 355-372, University of California Press, Berkeley, California, 1967.

Rubinstein, Reuven Y. and Ruth Marcus, "Efficiency of Multivariate Control Variates in Monte Carlo Simulation," *Operations Research*, vol. 33, pp. 661-677, May-June 1985.

Schatzoff, M., S. Fienberg, and R. Tsao, "Efficient Calculations of All-possible Regressions," *Technometrics*, vol. 10, pp. 768-779, 1968.

Schwarz, G., "Estimating the Dimension of a Model," *Annals of Statistics*, vol. 6, pp. 461-464, 1978.

Seber, G. A. F., in *Linear Regression Analysis*, John Wiley and Sons, New York, New York, 1977.

Seber, G. A. F., *Multivariate Observations,* John Wiley and Sons, New York, New York, 1984.

Siotani, M., T. Hayakawa, and Y. Fujikoshi, *Modern Multivariate Statistical Analysis: A Graduate Course and Handbook,* American Sciences Press, Columbus, Ohio, 1985.

Solberg, James J., CAN-Q User's Manual, School of Industrial Engineering, Purdue University West Lafayette, Indiana, 1980.

Thompson, Mary L., "Selection of Variables in Multiple Regression: A Review and Evaluation," *International Statistical Review*, vol. 46, pp. 1-19, 1978.

Venkatraman, Sekhar, Application of the Control Variate Technique to multiple Simulation Output Analysis, Department of Mechanical Engineering, The University of Texas, Austin, Texas, 1983.

Venkatraman, Sekhar and James R. Wilson, "The Efficiency of Control Variates in Multiresponse Simulation," *O.R. Letters*, vol. 5, no. 1, pp. 37-42, 1986.

Webster, J. T., R. F. Gunst, and R. L. Mason, "Latent Root Regression Analysis," *Technometrics*, vol. 16, pp. 513-522, 1974.

Welch, Peter D., "The Statistical Analysis of Simulation Results," in *Computer Preformance Modeling Handbook*, Academic Press Inc., New York, New York, 1983.

Wilson, J. R. and A. A. B. Pritsker, "Variance Reduction in Queueing Simulation using Generalized Concomitant Variables," *Journal of Statistical Computation and Simulation*, vol. 19, pp. 129-153, 1984a.

Wilson, J. R. and A. A. B. Pritsker, "Experimental Evaluation of Variance Reduction Techniques for Queueing Simulation using Generalized Concominant Variables," *Management Science*, vol. 30, pp. 1459-1472, Dec 1984b.

Wilson, James R., Variance Reduction Techiques for the Simulation of Queueing Networks, Technical Report , Mechanical Engineering Department, University of Texas, Austin, Texas, 1982.

Wilson, James R., "Variance Reduction Techniques for Digital Simulation," *American Journal of Mathematical and Management Sciences*, vol. 1, pp. 227-312, 1984 .

1 2

CONTROL VARIATE SELECTION FOR MULTIRESPONSE SIMULATION

A Thesis

Submitted to the Faculty

of

Purdue University

by

Kenneth William Bauer, Jr.

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

May 1987

ABSTRACT

Bauer, Kenneth W., Jr. Ph.D., Purdue University, May 1987.
Control Variate Selection for Multiresponse Simulation.
Major Professor: James R. Wilson.

A solution is offered to the general problem of optimal selection of control variates. Solutions are offered for two different cases of the general problem: (a) when the covariance matrix of the controls is unknown, and (b) when the covariance matrix of the controls is known and is incorporated into point and confidence region estimators. For the second case a new estimator is introduced. Under the assumption that the responses and the controls are jointly normal, the unbiasness of this new estimator is established , and its dispersion matrix is derived. A selection algorithm is implemented which locates the optimal subset of controls. The algorithm is based on criteria derived for the two cases listed above. A promising new class of controls is introduced which are called "routing variables". The asymptotic distribution of these controls is derived as well as their asymptotic mean and variance. Finally, the performance of the selection algorithm is investigated and the new estimator is contrasted with the classical estimator.

BIBLIOGRAPHY

Aitkin, M. A., "Simultaneous Inference and the Choice of Variable Subsets in Multiple Regression," *Technometrics*, vol. 16, pp. 221-227, 1974.

Akaike, H., "Information Theory and an Extension of the Maximum Likelihood Principle," *2nd International Symposium on Information Theory*, pp. 267-281, Akademiai Kiado, Budapest, 1973.

Allen, D. M., "Mean Square Error of Prediction as a Criterion for Selecting Variables," *Technometrics*, vol. 13, pp. 469-475, 1971.

Anderson, T. W., *An Introduction to Multivariate Statistical Analysis,* John Wiley, New York, New York, 1984.

Arvensen, J. N., "Jackknifing U-Statistics," *Annals of Mathematical Statistics*, vol. 40, pp. 2076-2100, 1969.

Bauer, Kenneth W., A Monte Carlo Study of Dimensionality Assessment and Factor Interpretation in Principal Components Analysis, Unpublished Masters Thesis, Air Force Institute of Technology, 1981.

Beaton, A. E., The Use of Special Matrix Operators in Statistical Calculus, Research Bulletin RB-64-51, Educational Testing Service, Princeton, New Jersey, 1964.

Box, George E., William G. Hunter, and J. Stuart, *Statistics for Experimenters,* John Wiley and Sons, New York, New York, 1978.

Cheng, R. C. H., "Analysis of Simulation Experiments under Normality Assumptions," *Journal of the Operational Research Society*, vol. 29, pp. 493-497, 1978.

Chvatal, Vasek, *Linear Programming,* W.H. Freeman and Company, New York, New York, 1980.

Crane, M. and A. J. Lemoine, "An Introduction to the Regenerative Method for Simulation Analysis," in *Lecture Notes in Control and Information Sciences*, Springer-Verlag, Berlin, Germany, 1977.

Draper, N. and H. Smith, *Applied Regression Analysis, Second Edition,* John Wiley and Sons, New York, New York, 1981.

Eakle, J. D., Regenerative Analysis Using Internal Controls, Unpublished Ph.D. Dissertation, Mechanical Engineering Department, University of Texas, Austin, Texas, 1982.

Flury, Bernard and Hans Reidwyl, "T ** 2 Tests and the Linear Two-Group Discriminant Function, and their Computation by Linear Regression," *The American Statistician*, vol. 39, pp. 20-25, February, 1985.

Furnival, G. M. and R. W. Wilson, "Regression by Leaps and Bounds," *Technometrics*, no. 16, pp. 499-511, 1974.

Furnival, George M., "All Possible Regressions with Less Computation," *Technometrics*, vol. 13, pp. 403-408, May 1971.

Gabriel, K. R., "Simultaneous Test Procedures - Some Theory of Multiple Comparisons," *Annals of Mathematical Statistics*, vol. 40, pp. 224-250, 1969.

Hocking, R. R., "The Analysis and Selection of Variables in Linear Regression," *Biometrics*, vol. 32, pp. 1-49, March, 1976.

Hocking, R. R., "Developments in Linear Regression Methodology: 1959-1982," *Technometrics*, vol. 25, pp. 219-230, August, 1983.

Hoerl, A. E. and R. W. Kennard, "Ridge Regression: Biased estimation for non-orthogonal problems," *Technometrics*, vol. 12, pp. 55-68, 1970.

Hogg, R. V. and A. T. Craig, *Introduction to Mathematical Statistics,* Macmillan Company, London, England, 1970.

Iglehart, D. L., "The Rengerative Method for Simulation Analysis," in *Current Trends in Programming Methodology, Vol III* , ed. K. M. Chandy and R. Yeh , Prentice Hall, Englewood Cliffs, New Jersey, 1978.

Iglehart, D. L. and P. A. W. Lewis, "Regenerative Simulation with Internal Controls," *Journal of the Association of Computing Machinery*, vol. 26, no. 2, pp. 271-282, 1979.

Johnson, R. A. and D. W. Wichern, *Applied Multivariate Analysis,* Prentice Hall, Inc. , Englewood Cliffs, New Jersey, 1982.

Jollife, I. T., "Disgarding Variables in Principal Components Analysis: Part I , Artifical Data," *Applied Statistics*, vol. 21, pp. 160-173, 1972.

Kennedy, W. J. and J. E. Gentle, *Statistical Computing,* Marcel Dekker, Inc., New York, New York, 1980.

Kenney, J. F. and E. S. Keeping, *Mathematics of Statistics Part II,* Van Nostrand Co. Inc., New York, New York, 1951.

Kleijnen, J. P. C., *Statistical Techniques in Simulation, Part 1 and 2,* Marcel Deckker, New York, New York, 1975.

Lavenberg, S. S., T. L. Moeller, and P. D. Welch, Statistical Results on Multiple Control Variables with Application to Variance Reduction in Queueing Network Simulation, IBM Research Report RC-7423, Yorktown Heights, New York, 1978.

Lavenberg, S. S., T. L. Moeller, and C. H. Sauer, "Concomitant Control Variables Applied to the Regenerative Simulation of Queueing Systems," *Operations Research*, vol. 27, no. 1, pp. 134-160, 1979.

Lavenberg, S. S. and P. D. Welch, "A Perpective on the Use of Control Variables to Increase the Efficiency of Monte Carlo Simulations," *Management Sciences*, vol. 27, pp. 322-334, March 1981.

Lavenberg, S. S., T. L. Moeller, and P. D. Welch, "Statistical Results on Control Variables with Application to Queueing Network Simulation," *Operations Research*, vol. 30, pp. 182-202, Jan-Feb 1982.

Lavenberg, Stephen S., T. L. Moeller, and P. D. Welch, "Statistical Results on Control Variables with Applications to Queueing Network Simulation," *Operation Research*, vol. 30, pp. 182-202.

Lindley, D. V., "The Choice of Variables in Multiple Regression," *Journal of the Royal Statistical Society*, vol. B30, pp. 31-53, 1968.

Mallows, C. L., "Some Comments on C sub p," *Technometrics*, vol. 15, pp. 661-675, 1973.

Mantel, N., "Why Stepdown Procedures in Variable Selection," *Technometrics*, vol. 12, pp. 621-625, 1970.

McCabe, George P., "Evaluation of Regression Coefficient Estimates Using alpha-acceptabilty," *Technometrics*, vol. 20, pp. 131-139, May 1978.

McCabe, George P., "Principal Variables," *Technometrics*, vol. 26, pp. 137-144, May, 1984.

McKay, R. J., "Variable Selection in Multivariate Regression: An Application of Simultaneous Test Procedures," *Journal of the Royal Statistical Society*, vol. B 39, pp. 371-380, 1977.

Miller, Rupert G., "The jackknife - a review," *Biometrika*, vol. 61, pp. 1-15, 1974.

Muirhead, R. J., *Aspects of Multivariate Statistical Theory,* John Wiley and Sons, New York, New York, 1982.

Neter, J., W. Wasserman, and M. H. Knuter, *Applied Linear Regression Models,* Richard D. Erwin, Inc., Homewood, Illinois, 1983.

Neuts, Marcel F., *Probability,* Allyn and Bacon, Inc., Boston, Mass., 1973.

Nova, A. M. Porta, A Generalized Approach to Variance Reduction in Discrete-event Simulation using Control Variables, Unpublished Ph.D. Dissertation, Department of Mechanical Engineering, The University of Texas, Austin, Texas, 1985.

Nozari, A., S. F. Arnold, and C. D. Pegden, "Control Variates for Multipopulation Experiments," *IIE Transactions*, vol. 16, pp. 159-169, June, 1984.

Pritsker, A. Alan B., *Introduction to Simulation and SLAM,* Halsted Press, New York, New York, 1986.

Rao, C. R., "Least Squares Theory using an Estimated Dispersion Matrix and its Application to Measurement of Signals," *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probabilty,* vol. I, pp. 355-372, University of California Press, Berkeley, California, 1967.

Rubinstein, Reuven Y. and Ruth Marcus, "Efficiency of Multivariate Control Variates in Monte Carlo Simulation," *Operations Research*, vol. 33, pp. 661-677, May-June 1985.

Schatzoff, M., S. Fienberg, and R. Tsao, "Efficient Calculations of All-possible Regressions," *Technometrics*, vol. 10, pp. 768-779, 1968.

Schwarz, G., "Estimating the Dimension of a Model," *Annals of Statistics*, vol. 6, pp. 461-464, 1978.

Seber, G. A. F., in *Linear Regression Analysis*, John Wiley and Sons, New York, New York, 1977.

Seber, G. A. F., *Multivariate Observations,* John Wiley and Sons, New York, New York, 1984.

Siotani, M., T. Hayakawa, and Y. Fujikoshi, *Modern Multivariate Statistical Analysis: A Graduate Course and Handbook,* American Sciences Press, Columbus, Ohio, 1985.

Solberg, James J., CAN-Q User's Manual, School of Industrial Engineering, Purdue University West Lafayette, Indiana, 1980.

Thompson, Mary L., "Selection of Variables in Multiple Regression: A Review and Evaluation," *International Statistical Review,* vol. 46, pp. 1-19, 1978.

Venkatraman, Sekhar, Application of the Control Variate Technique to multiple Simulation Output Analysis, Department of Mechanical Engineering, The University of Texas, Austin, Texas, 1983.

Venkatraman, Sekhar and James R. Wilson, "The Efficiency of Control Variates in Multiresponse Simulation," *O.R. Letters,* vol. 5, no. 1, pp. 37-42, 1986.

Webster, J. T., R. F. Gunst, and R. L. Mason, "Latent Root Regression Analysis," *Technometrics,* vol. 16, pp. 513-522, 1974.

Welch, Peter D., "The Statistical Analysis of Simulation Results," in *Computer Preformance Modeling Handbook*, Academic Press Inc., New York, New York, 1983.

Wilson, J. R. and A. A. B. Pritsker, "Variance Reduction in Queueing Simulation using Generalized Concomitant Variables," *Journal of Statistical Computation and Simulation,* vol. 19, pp. 129-153, 1984a.

Wilson, J. R. and A. A. B. Pritsker, "Experimental Evaluation of Variance Reduction Techniques for Queueing Simulation using Generalized Concominant Variables," *Management Science,* vol. 30, pp. 1459-1472, Dec 1984b.

Wilson, James R., Variance Reduction Techiques for the Simulation of Queueing Networks, Technical Report , Mechanical Engineering Department, University of Texas, Austin, Texas, 1982.

Wilson, James R., "Variance Reduction Techniques for Digital Simulation," *American Journal of Mathematical and Management Sciences,* vol. 1, pp. 227-312, 1984 .

PURDUE UNIVERSITY

Graduate School

This is to certify that the thesis prepared

By Kenneth W. Bauer, Jr.

Entitled

Control Variate Selection for Multiresponse Simulation

Complies with University regulations and meets the standards of the Graduate School for originality and quality

For the degree of Doctor of Philosophy

Signed by the final examining committee:

James R. Wilson, chair

Arnold L. Sweet

J. Talavage

George McCabe

Approved by the head of school or department:

April 28 1987

This thesis ☐ is ☒ is not to be regarded as confidential

James R. Wilson

Major professor

Grad. School
Form No. 9
Revised 9-85

To Cindy

Love is the answer to all questions.

-- St. Paul --

Little Darling
Its been a long, cold lonely winter
Little Darling
It seems like years since its been here
Here comes the sun
Here comes the sun
And I say
Its alright

-- George Harrison --

To Scott and Steve, also

Elfin magic and Cowboy dust,
saviors two from childhood's end.

ACKNOWLEDGEMENTS

I acknowledge the good Lord, who with frightening precision, gave me the requisite intelligence and creativity to accomplish this dissertation. The marvelous clay of God, we mold ourselves. I thank the United States Air Force for sponsoring my study here at Purdue. I greatly appreciate living in a country where even a former janitor can aspire to and win a doctorate.

I would like to thank my advisor, Dr. James R. Wilson, for his critical eye and high standards. I am proud to have been his student. I acknowledge the help of my committee: Dr. George McCabe, Dr. Arnold Sweet, and Dr. Joseph Talavage, Full Professors in the finest tradition.

I acknowledge Sekhar Venkatraman, a friend and colleague. I give special thanks to Moses Sudit for helping me in the formatting of this document. I am honored to have made such a fine friend. Finally, I would acknowledge Dr. Leon Gleser: the toughest, fairest, most unforgettable personality of my college years.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

ABSTRACT

Bauer, Kenneth W., Jr. Ph.D., Purdue University, May 1987.
Control Variate Selection for Multiresponse Simulation.
Major Professor: James R. Wilson.

A solution is offered to the general problem of optimal selection of control variates. Solutions are offered for two different cases of the general problem: (a) when the covariance matrix of the controls is unknown, and (b) when the covariance matrix of the controls is known and is incorporated into point and confidence region estimators. For the second case a new estimator is introduced. Under the assumption that the responses and the controls re jointly normal, the unbiasness of this new estimator is established , and its dispersion matrix is derived. A selection algorithm is implemented which locates the optimal subset of controls. The algorithm is based on criteria derived for the two cases listed above. A promising new class of controls is introduced which are called "routing variables". The asymptotic distribution of these controls is derived as well as their asymptotic mean and variance. Finally, the performance of the selection algorithm is investigated and the new estimator is contrasted with the classical estimator.

CHAPTER 1
INTRODUCTION

The method of control variables is one of the main variance reduction techniques used in discrete event simulation. This method attempts to exploit correlations between output responses and associated auxiliary variables with known means that can be observed during the course of a simulation run. Although control variables can be *external* (that is, similar variables in a much simplified version of the original model which is driven by the same random number streams as the original model), our research deals only with *internal* or so-called *concomitant* controls.

There are several tactical issues that must be addressed to employ control variables successfully. One such issue is efficiency. Several authors (see review of the literature) have addressed the fact that a trade-off must be recognized in the application of the control variable technique. Various loss factors have been derived to quantify the diminishing marginal returns that are experienced (on the average) when additional control variables are included in the variance reduction scheme. This trade-off arises because the application of control variables requires the estimation of additional control coefficients. If the sample size is taken to be constant, then the variance reduction achieved by the use of additional controls can be offset by the variance inflation due to the estimation of additional coefficients. Hence a

selection scheme is needed to pick a good subset of the candidate controls.

The control variable selection problem is important because more often than not a simulator trying to use control variables finds himself confronted with multiple candidates for controls. The literature to date only offers ad hoc methods to solve this problem. Several authors have called for research into this problem; in particular Lavenberg, Moeller, and Welch (1982), Rubinstein and Marcus (1985), and Venkatraman and Wilson (1986) have all suggested methods for developing an effective control variate selection procedure. Unfortunately there has been no follow-up work on any of these proposals.

1.1 Research Objectives

Our primary objective is to formulate and evaluate control variate selection criteria for multiresponse simulation experiments in which we seek point and confidence region estimators for the mean response. We distinguish the following cases: (a) the covariance matrix of controls is unknown, and (b) the covariance matrix of the controls is known and is incorporated into the point and confidence region estimator. The second case requires the introduction of a new point estimator and the the derivation of its mean vector and covariance matrix. We also introduce a new class of controls that we call "routing variables", and we establish the asymptotic distribution of these controls so that we may exploit them not only in case (a) but also in case (b). The experimental evaluation phase of the research includes a comparison of the performance of the controlled estimation procedures described in (a) and (b) above.

1.2 Organization of the Research

We review both the control variable literature as well as the pertinent statistical literature which bears on variable selection in the context of linear regression. The literature review is presented in Chapter 2. Chapter 3 presents theoretical arguments which lead to selection criteria for both cases mentioned above. In Chapter 3 we also derive the properties of the new estimator $\bar{Y}(\hat{\gamma})$. Chapter 4 describes our experimental setup. We discuss the models used, the experimental layout, and the necessary matrix methods required to implement the selection algorithm. In Chapter 5 we summarize the results of our experiments. In Chapter 6 we present an overview of the research and propose directions for future research.

CHAPTER 2
LITERATURE REVIEW

The first notable, comprehensive discussion of control variables (actually variance reduction techniques in general) is offered by Kleijnen (1974). A more rigorous, up to date survey of variance reduction techniques is found in Wilson (1984).

2.1 Univariate Simulation Response with a Single Control

Assume Y is an estimator of μ_Y, where μ_Y is the mean of some response of interest. Let X be a variable observed during the course of the simulation. We assume that X is highly correlated with the response, and further that its mean μ_X is known. The variable X is the control variable.

Consider the "controlled estimator"

$$Y(b) = Y - b(X - \mu_X) \tag{2.1.1}$$

Note, if b is a constant

$$E(Y(b)) = \mu_Y , \tag{2.1.2}$$

$$\text{and} \quad \text{var}(Y(b)) = \text{var}(Y) + b^2\text{var}(X) - 2b\,\text{cov}(Y,X) \,. \tag{2.1.3}$$

So Y(b) is an unbiased estimator of μ_Y. The variance of Y(b) will be smaller than the variance of Y if

$$2b\,\text{cov}(Y,X) > b^2\text{var}(X) \,. \tag{2.1.4}$$

A little calculus reveals that

$$\beta = \frac{\text{cov}(Y,X)}{\text{var}(X)} \tag{2.1.5}$$

minimizes (2.1.3). Plugging (2.1.5) into (2.1.3) yields the minimum variance

$$\text{var}(Y(\beta)) = (1-\rho_{XY}^2)\,\text{var}(Y)\,, \tag{2.1.6}$$

where ρ_{YX} is the correlation coefficient between Y and X. Following Porta Nova (1985), we obtain an unbiased point estimator of μ_Y by averaging the controlled observations

$$Y_i(\beta) = Y_i - \beta(X_i - \mu_X), \quad i = 1, \ldots, K, \tag{2.1.7}$$

to form

$$\bar{Y}(\beta) = \sum_{i=1}^{K} Y_i(\beta)/K, \tag{2.1.8}$$

where K is the sample size. Since we do not know the optimal value β, we must estimate it. An intuitive estimate of β replaces the right-hand side of (2.1.5) with the appropriate sample quantities. This solution turns out to be the least squares solution for β. When the assumption of joint normality between Y and X is made, then the least squares solution is also the maximum likelihood solution. We estimate β by

$$\hat{\beta} = \frac{\sum_{i=1}^{K}(Y_i - \overline{Y})(X_i - \overline{X})}{\sum_{i=1}^{K}(X_i - \overline{X})^2}, \tag{2.1.9}$$

and the point estimator of μ_Y is then

$$\hat{\mu}_Y(\hat{\beta}) = \sum_{i=1}^{K} Y_i(\hat{\beta})/K. \tag{2.1.10}$$

We obtain an interval estimate for μ_Y by application of regression theory. First we make note of what happens under the assumption of joint normality for Y and X. In this situation the conditional distribution of Y given X is also normal:

$$Y \mid X{=}x \sim N(\mu_Y + \beta(x - \mu_X), \sigma_\epsilon^2) \tag{2.1.11}$$

where

$$\sigma_\epsilon^2 = \sigma_Y^2(1 - \rho_{XY}^2) \tag{2.1.12}$$

and

$$\sigma_Y^2 = \text{var}(Y) \tag{2.1.13}$$

We see that if X=x, there is a linear regression of Y on X. Given we know values of the control variable X, as well as its mean, we see that the conditional mean of Y has two terms. The first term is μ_Y, the parameter to be estimated. The second term is a correction due to the particular values of the control. To get at μ_Y, we will subtract out these corrections as in (2.1.7). Equation (2.1.11) shows us that each observed Y_i has the form

$$Y_i = \mu_Y + \beta(X_i - \mu_X) + \epsilon_i \quad 1 \leq i \leq K, \tag{2.1.14}$$

where ϵ_i are the residuals

$$\epsilon_i \sim N(0, \sigma_\epsilon^2) \tag{2.1.15}$$

There are two unknown quantities in (14), so we can apply the method of least squares to solving for μ_Y and β. The parameter μ_Y is the intercept of equation (14) and under the joint normality assumption for X and Y,

$$\hat{\mu}_Y(\hat{\beta}) \sim N(\mu_Y, \sigma_\epsilon^2 s_{11}), \tag{2.1.16}$$

where s_{11} is the upper left-hand corner entry of the matrix $(\mathbf{D}'\mathbf{D})^{-1}$ where

$$\mathbf{D} = \begin{bmatrix} 1 & X_1 - \mu_X \\ 1 & X_2 - \mu_X \\ 1 & X_3 - \mu_X \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & X_K - \mu_X \end{bmatrix}. \tag{2.1.17}$$

Now to form a confidence interval about $\hat{\mu}_Y(\hat{\beta})$ we will first need an estimate of σ_ϵ^2. Remembering that σ_ϵ^2 represents that variability in Y given we have accounted for X, the formula for the residual mean square error given in regression theory makes good intuitive sense as an estimator of σ_ϵ^2, that is

$$\hat{\sigma}_\epsilon^2 = \frac{\sum_{i=1}^{K}(Y_i - \hat{Y}_i)^2}{K-2}, \tag{2.1.18}$$

where

$$\hat{Y}_i(\hat{\beta}) = \hat{\mu}_Y(\hat{\beta}) + \hat{\beta}(X_i - \mu_X), \quad 1 \le i \le K \tag{2.1.19}$$

Now it can be shown (Hogg and Craig (1970), pg. 337) that

$$\frac{\hat{\mu}_Y(\hat{\beta}) - \mu_Y}{\left[\dfrac{\hat{\sigma}_\epsilon^2 s_{11}}{K-2}\right]^{1/2}} \sim t_{K-2}, \tag{2.1.20}$$

where t_{K-2} is a Student-t distribution with K-2 degrees of freedom. From regression theory s_{11} is given by (Draper and Smith (1981), pg. 83 and some algebra because our X_i is $X_i - \mu_X$)

$$s_{11} = \frac{\sum_{i=1}^{K}(X_i - \mu_X)^2}{K \sum_{i=1}^{K}(X_i - \bar{X})^2}, \tag{2.1.21}$$

where

$$\bar{X} = \frac{\sum_{i=1}^{K} X_i}{K}. \tag{2.1.22}$$

A 100(1-α)% confidence interval is given by

$$\hat{\mu}_Y(\hat{\beta}) \pm t_{K-2}(1-\frac{\alpha}{2})\hat{\sigma}_\epsilon\sqrt{s_{11}}. \tag{2.1.23}$$

As mentioned in the introduction we expect to incur a loss due to the estimation of β. In this case where we only have a single control variable, we expect that the realized variance reduction should, on the average, decrease as sample size decreases. This loss is quantified via the loss factor. Following Lavenberg, Moeller and Welch (1982), we define the loss factor as the ratio of the variance of the estimator of μ_Y when the optimal control coefficient is not known to the the variance of the estimator when the coefficient is known. So

$$LF = \frac{\mathrm{var}(\hat{\mu}_Y(\hat{\beta}))}{\mathrm{var}(\hat{\mu}_Y(\beta))} = \frac{\mathrm{var}(\overline{Y}(\hat{\beta}))}{\mathrm{var}(\overline{Y}(\beta))}\ . \tag{2.1.24}$$

In the next section we give details on the derivation of the loss factor when there are more than one control variable. The loss factor here is a special case (Q=1) and hence from equation (2.2.32)

$$LF = \frac{K-2}{K-3}. \tag{2.1.25}$$

The loss factor acts as a multiplier to the minimum variance ratio (MVR) where

$$MVR = \frac{\mathrm{var}(\overline{Y}(\beta))}{\mathrm{var}(\overline{Y})} \tag{2.1.26}$$

which represents the variance reduction achievable when the optimal control coefficients are known. Multiplying the loss factor by the minimum variance ratio allows for the loss in potential variance reduction due to the estimation of β. This product is the variance ratio (VR) and

$$VR = LF \times MVR \tag{2.1.27}$$

Later we will change the abbreviations to more standard Greek symbols.

2.2 Univariate Simulation Response with Multiple Controls

The previous discussion can be extended to the case of multiple controls. We summarize the development presented by Lavenberg and Welch (1981) for simulation output analysis based on independent replications, batch means, and regenerative analysis.

2.2.1 Output Analysis Using Independent Replications or Batch Means

During the course of making simulation runs, we observe the values of the response of interest as well as the Q control variables. Separate observations could occur as the result of independent replications of the simulation model. These observations could also result from the use of batching to form nearly independent observations. Let $\mathbf{X}$ be a $Q \times 1$ vector of controls, i.e., $\mathbf{X} = (X_1, \ldots, X_Q)'$ with known mean vector $\mu_{\mathrm{X}} = (\mu_1, \ldots, \mu_Q)'$ and let $B = (b_1, \ldots, b_Q)$ be a $1 \times Q$ row vector of constants, then the controlled estimator of μ_Y becomes

$$Y(B) = Y - B(\mathbf{X} - \mu_{\mathrm{X}}) \,. \tag{2.2.1}$$

The vector β which minimizes Var(Y(B)) is given by

$$\beta = \sigma_{Y\mathrm{X}} \Sigma_{\mathrm{XX}}^{-1}, \tag{2.2.2}$$

where Σ_{XX} is the $Q \times Q$ covariance matrix of the controls and $\sigma_{Y\mathrm{X}}$ is the $1 \times Q$ vector of covariances between the response and the controls. See Anderson

(1984), pg. 39, for a proof. The resulting minimum variance is

$$Var(Y(\beta)) = (1 - \rho^2_{Y\mathbf{X}})\ \mathrm{var}(Y), \tag{2.2.3}$$

where $\rho_{Y\mathbf{X}}$ is *the coefficient of multiple correlation* between Y and **X**. The authors next comment on the availability and choice of control variables. They cite many application papers and distinguish between external and concomitant controls.

The previous discussion hinged on the assumption that β is known. This, of course, is not the case in practice (otherwise there would be little need to simulate a process in the first place). We must estimate β and incorporate the estimate into an effective statistical procedure to estimate μ_Y. To obtain an unbiased estimator of μ_Y, we make K independent replications of the model or we organize the output from one run into K batches so that means computed from each batch are approximately independent. If $\mathbf{X}_k$ is the vector of controls, observed on the k^{th} replication or batch, then we compute

$$Y_k(B) = Y - B(\mathbf{X}_k - \mu_k)\ ,\ k = 1, \ldots ,K. \tag{2.2.4}$$

A sensible estimator from the entire data set would be

$$\overline{Y}(B) = \frac{1}{K} \sum_{k=1}^{K} Y_k(B)\ . \tag{2.2.5}$$

Now

$$\mathrm{var}(\overline{Y}(B)) = \frac{1}{K}\ \mathrm{var}(Y(B))\ . \tag{2.2.6}$$

However, we still do not know the optimal value of B, and we must estimate it. One estimate of β is

$$\hat{\beta} = \hat{\sigma}_{YX}\hat{\Sigma}_{XX}^{-1}, \tag{2.2.7}$$

where $\hat{\Sigma}_{XX}$ and $\hat{\sigma}_{YX}$ are the sample analogs of Σ_{XX} and σ_{YX}. We now substitute $\hat{\beta}$ for B in (2.2.4) and (2.2.5) to produce the estimates $Y_k(\hat{\beta})$ and $\bar{Y}(\hat{\beta})$. In general $\bar{Y}(\hat{\beta})$ is not an unbiased estimator because $\hat{\beta}$ and $\bar{\mathbf{X}}$ are not in general independent. A simplfying assumption is that $(Y, \mathbf{X})$ are jointly distributed as multivariate normal random variables, see Cheng (1978). This assumption may be justified by the use of sample means as controls (as well as the estimator Y). In this case $\bar{Y}(\hat{\beta})$ is an unbiased estimator of μ_Y, and using regression theory (discussed in greater detail later) we obtain an estimator of $\text{var}(\bar{Y}(\hat{\beta}))$ such that

$$\frac{\bar{Y}(\hat{\beta}) - \mu_Y}{\hat{\text{var}}(\bar{Y}(\hat{\beta}))} \sim t_{K-Q-1}, \tag{2.2.8}$$

where t_{K-Q-1} is a t-distributed random variable with K-Q-1 degrees of freedom. Q is the number of controls. Explicitly

$$\hat{\text{var}}(\bar{Y}(\hat{\beta})) = \hat{\sigma}_c^2 s_{11},$$

where $\hat{\sigma}_c^2$ is given by (2.2.38) and s_{11} is described in (2.1.17). This leads to the familiar $100(1-\alpha)\%$ confidence interval for μ_Y

$$\bar{Y}(\hat{\beta}) \pm t_{K-Q-1}(1-\alpha/2)\sqrt{s_{11}}\hat{\sigma}_c \tag{2.2.9}$$

The procedure used to construct the interval given in (2.2.9) is discussed in greater detail in a subsequent section.

Since β is estimated by $\hat{\beta}$, one would expect that some loss in variance reduction would be incurred. We define the loss factor to be the ratio of the variance of the controlled estimator when the controls are unknown (hence must be estimated) to the variance of the controlled estimator when the controls are known. Using the notation of Venkatraman and Wilson (1986), we let λ_1 denote the loss factor, we will show

$$\lambda_1 = \frac{K-2}{K-Q-2} \ . \tag{2.2.10}$$

This factor is derived from the following considerations. If we do not use controls then by direct estimation

$$\mathrm{var}(\overline{Y}) = \frac{\sigma_Y^2}{K} \ , \tag{2.2.11}$$

where σ_Y^2 is the variance of Y. Now if we know the control coefficients, then from (2.2.3)

$$\mathrm{var}(\overline{\mathbf{Y}}(\hat{\gamma})) = (1-\rho_{Y\mathbf{X}}^2)\frac{\sigma_Y^2}{K} \tag{2.2.12}$$

The ratio of (2.2.12) and (2.2.11) is called the minimum variance ratio (η_1^*)

$$\eta_1^* \equiv \frac{\mathrm{var}(\overline{\mathbf{Y}}(\hat{\gamma}))}{\mathrm{var}(\overline{Y})} = 1-\rho_{Y\mathbf{X}}^2, \tag{2.2.13}$$

and we see that $100(1-\eta_1^*)$ is the percentage variance reduction achievable when B is known.

When we estimate β with $\hat{\beta}$ we are now interested in

$$\eta_1 \equiv \frac{\text{var}(\overline{Y}(\hat{\beta}))}{\text{var}(\overline{Y})} \tag{2.2.14}$$

η_1 is called the variance ratio. We note

$$\eta_1 \equiv \frac{\text{var}(\overline{Y}(\hat{\beta}))}{\text{var}(\overline{Y}(\beta))} \frac{\text{var}(\overline{Y}(\beta))}{\text{var}(\overline{Y})} = \lambda_1 \eta_1^* , \tag{2.2.15}$$

where λ_1 is the loss factor due to the estimation of B.

At this point we have all the pieces save an expression for $\text{var}(\overline{Y}(\hat{\beta}))$. The following details are from Lavenberg, Moeller, and Welch (1982). First we remember (2.2.5)

$$\overline{Y}(\hat{\beta}) = \overline{Y} - \hat{\beta}(\overline{\mathbf{X}} - \mu_{\mathbf{X}}). \tag{2.2.16}$$

To get $\text{var}(\overline{Y}(\hat{\beta}))$ the technique will be as follows. First we will write $\text{var}(\overline{Y}(\hat{\beta}))$ as a linear combination of the Y_k, then we will fix the controls, compute the conditional variance, and finally, exploit the conditional unbiasedness of $\overline{Y}(\hat{\beta})$ by computing the variance of $\overline{Y}(\hat{\beta})$ as the expected value (with respect to the controls) of the conditional variance.

Define M as a $Q \times K$ matrix such that

$$\mathrm{M} = \begin{bmatrix} x_{11}-\bar{x}_1 & \dots & x_{1K}-\bar{x}_1 \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ x_{Q1}-\bar{x}_Q & \dots & x_{QK}-\bar{x}_Q \end{bmatrix}, \tag{2.2.17}$$

where x_{ij} is the value of the i^{th} control on the j^{th} replication or batch, also $\bar{x}_i$ is the sample mean of the i^{th} control. From (2.2.2) we can write $\hat{B}$ as

$$\hat{\beta} = \hat{\sigma}_{Y\mathbf{X}}\hat{\Sigma}_{\mathbf{XX}}^{-1} = (Y - \bar{Y}\mathbf{1}_K)'\mathrm{M}'(\mathrm{MM}')^{-1} \tag{2.2.18}$$

where $\mathbf{1}_K$ is a column vector of 1s. Now we can write

$$\bar{Y}(\hat{\beta}) = b'Y , \tag{2.2.19}$$

where

$$b' = \frac{1}{K}\mathbf{1}'_K - (\bar{\mathbf{X}} - \mu_{\mathbf{X}})'(\mathrm{MM}')^{-1}\mathrm{M} . \tag{2.2.20}$$

Given $\mathbf{X}_k = x_k$ for k=1, ... , K, then b' is a constant vector. Now we have the conditional estimator in terms of the Y_k. We compute

$$\mathrm{var}(\bar{Y}(\hat{\beta}) \mid \mathbf{X}_k = x_k \text{ for all } k) = b'\Big[(\mathrm{var}Y \mid \mathbf{X}_k = x_k \text{ for all } k)\Big]b , \tag{2.2.21}$$

which reduces to

$$\mathrm{var}(\overline{Y}(\hat{\beta}) \mid \mathbf{X}_k = x_k \text{ for all } k) = \sigma_\epsilon^2 \left[\frac{1}{K} + (\overline{\mathbf{X}} - \mu_{\mathbf{X}})'(\mathrm{MM}')^{-1}(\overline{\mathbf{X}} - \mu_{\mathbf{X}}) \right], \tag{2.2.22}$$

where σ_ϵ is the residual variance (as described in (2.2.34)). Now we find the expected value of (2.2.22) with respect to the controls, and we get

$$\mathrm{var}(\overline{Y}(\hat{\beta})) = E_{\mathbf{X}} \left(\frac{\sigma_\epsilon^2}{K} \left[1 + K(\overline{\mathbf{X}} - \mu_{\mathbf{X}})'(\mathrm{MM}')^{-1}(\overline{\mathbf{X}} - \mu_{\mathbf{X}}) \right] \right). \tag{2.2.23}$$

Now since

$$\hat{\Sigma}_{\mathbf{X}} = \frac{\mathrm{MM}'}{K-1}, \tag{2.2.24}$$

we have

$$\mathrm{var}(\overline{Y}(\hat{\beta})) = E_{\mathbf{X}} \left(\frac{\sigma_\epsilon^2}{K} \left[1 + \frac{K}{K-1}(\overline{\mathbf{X}} - \mu_{\mathbf{X}})'\hat{\Sigma}_{\mathbf{X}}^{-1}(\overline{\mathbf{X}} - \mu_{\mathbf{X}}) \right] \right). \tag{2.2.25}$$

We note that (Anderson (1984))

$$T^2 \equiv K(\overline{\mathbf{X}} - \mu_{\mathbf{X}})\hat{\Sigma}_{\mathbf{X}}^{-1}(\overline{\mathbf{X}} - \mu_{\mathbf{X}}) \tag{2.2.26}$$

is Hotelling's T^2 statistic. Also Corollary 5.2.1 of Anderson (1984) gives

$$\frac{K(\bar{\mathbf{X}}-\mu_{\mathbf{X}})'\hat{\Sigma}_{\mathbf{X}}^{-1}(\bar{\mathbf{X}}-\mu_{\mathbf{X}})}{K-1}\,\frac{K-Q}{Q}\sim F_{Q,K-Q}\ . \tag{2.2.27}$$

Kenney and Keeping (1951) give

$$E(F_{Q,K-Q}) = \frac{K-Q}{K-Q-2}\ . \tag{2.2.28}$$

Now (2.2.25) becomes

$$\left(\frac{\sigma_\epsilon^2}{K}\left[1+\frac{1}{K-1}E(T^2)\right]\right) = \frac{\sigma_\epsilon^2}{K}\left(1+\frac{Q}{K-Q}E(F_{Q,K-Q})\right) \tag{2.2.29}$$

which finally reduces to

$$\mathrm{var}(\bar{Y}(\hat{\beta})) = \frac{\sigma_\epsilon^2}{K}\left(1+\frac{Q}{K-Q-2}\right). \tag{2.2.30}$$

Examination of (2.2.14) reveals that λ_1 , the loss factor, is

$$\lambda_1 \equiv \frac{\mathrm{var}(\bar{Y}(\hat{\beta}))}{\mathrm{var}(\bar{Y}(\beta))}\ , \tag{2.2.31}$$

so from (2.2.12) and (2.2.30)

$$\lambda_1 = \frac{K-2}{K-Q-2}\ . \tag{2.2.32}$$

Now that we have a theory which lets us develop confidence intervals and quantifies the loss incurred due to the estimation of β, we need a statistically valid procedure to construct the intervals. Lavenberg et al. develop procedures based on the method of independent replications as well as the regenerative method. Here we summarize only the method of independent replications, and in a later section we discuss procedures for the regenerative method.

Now, Y, $\mathbf{X} = (x_1, \ldots, x_Q)'$ are assumed to be jointly distributed as a multivariate normal. Conditional on $\mathbf{X} = x$, Y will be distributed as univariate normal with

$$E(Y \mid \mathbf{X} = x) = \mu_y + \beta(\mathbf{X} - \mu_{\mathrm{X}}) , \tag{2.2.33}$$

where β is given by (2.2.2) (the optimal control coefficient vector). The variance is given by

$$\mathrm{var}(Y \mid \mathbf{X} = x) = \sigma_Y^2 (1 - \rho_{Y\mathrm{X}}^2) . \tag{2.2.34}$$

So if we take the $\mathbf{X}$ as fixed we have the linear regression problem with

$$\begin{pmatrix} Y_1 \\ \cdot \\ \cdot \\ \cdot \\ Y_k \end{pmatrix} = \begin{pmatrix} 1 & x_{11} - \mu_{x_1} & \ldots & x_{Q1} - \mu_{x_Q} \\ & \cdot & & \cdot \\ & \cdot & & \cdot \\ & \cdot & & \cdot \\ 1 & x_{1K} - \mu_{x_1} & \ldots & x_{QK} - \mu_{x_Q} \end{pmatrix} \begin{pmatrix} \mu_Y \\ \beta_1 \\ \cdot \\ \cdot \\ \cdot \\ \beta_Q \end{pmatrix} + \epsilon, \tag{2.2.35}$$

where x_{ij} is as given in (2.2.17) and μ_{x_i} is the is the known mean of the i^{th} control. It becomes apparent that we will be estimating μ_Y with the least squares estimate for the intercept of (2.2.35). We form our confidence interval in the standard manner.

Let $\hat{\mu}_Y$ and $\hat{\beta}$ be the corresponding estimators of μ_Y and β and let $\mathbf{D}$ denote the $K\times(Q+1)$ matrix on the right hand side of (2.2.35). From regression theory the conditional distribution of $\hat{\beta}$ given $\mathbf{D}$ is

$$\hat{\beta} \sim N_{Q+1}(\beta,\, \sigma_\epsilon^2(\mathbf{D}'\mathbf{D})^{-1}) \,. \tag{2.2.36}$$

Hence

$$\hat{\mu}_Y \sim N(\mu_Y, \sigma_\epsilon^2 s_{11}) \,, \tag{2.2.37}$$

where s_{11} is the upper left most corner of $(\mathbf{D}'\mathbf{D})^{-1}$. Now all that is required is an estimate of the common variance σ_ϵ^2. Such an estimate is given by

$$\hat{\sigma}_\epsilon^2 = \frac{\left(\sum_{k=1}^{K} Y_k^2 - \sum_{k=1}^{K} (\hat{\mu}_Y + \hat{\beta}(x_k - \mu_x))^2\right)}{K-Q-1} \,. \tag{2.2.38}$$

So given the observed values of the control variables a $100(1-\alpha)$ % confidence interval for μ_Y is given by

$$\hat{\mu}_Y \pm t_{K-Q-1}(1-\alpha/2)\sqrt{s_{11}}\ \hat{\sigma}_\epsilon \,. \tag{2.2.39}$$

2.2.2 Output Analysis Using the Regenerative Method

Lavenberg and Welch (1981) summarize a methodology for the construction of confidence intervals based on the regenerative method. A more detailed development is found in Lavenberg, Moeller and Sauer (1979) as well as in Iglehart and Lewis (1979).

The regenerative method can be based on a single run of the model. The method may be applied if there exists an increasing sequence of random times that partition a run into independent and identically distributed cycles. This sequence of regeneration times typically correspond to some distinguished state of the model. This state is such that, when it is entered, the model starts afresh according to the same probabilistic mechanism that drove the previous cycles. Lavenberg and Welch (1981) point out that, in complex simulations, regeneration points may occur so infrequently as to discourage the use of this method. The construction of confidence intervals using the regenerative method is discussed in detail in Crane and Lemoine (1977), Iglehart (1978) and Welch (1983).

We follow Iglehart and Lewis (1979) in their application of the regenerative method using control variables. Assume we observe in a run of n cycles:

$$(Y_j, \tau_j, \mathbf{X}_j): \ 1 \leq j \leq n \ ,$$

where τ_j is the length of the j^{th} cycle, Y_j is some response of interest and $\mathbf{X}_j$ is a $Q \times 1$ vector of controls. Many steady-state parameters of interest can be expressed as the ratio of two expected values. Let r be such a variable. A

"controlled" point estimator of

$$r = \frac{E[Y]}{E[\tau]}$$

is given by

$$\hat{r}(b) = \frac{\left[\overline{Y} - b\overline{\mathbf{X}}\right]}{\overline{\tau}} \tag{2.2.40}$$

where b is a $1 \times Q$ row vector of control coefficients and $\overline{Y}$, $\overline{\mathbf{X}}$ and $\overline{\tau}$ are the sample means of Y, $\mathbf{X}$ and τ. We note that controls are only being applied to the numerator of the estimator. This type of estimator is called a top-controlled estimator. Eakle (1982) developed a two-stage procedure which first applied controls to the denominator of (2.2.40) to reduce the bias of $\hat{r}$ and then applied another set of control variables to the numerator to reduce the variance of the estimator. We discuss only top-controlled estimators.

To obtain an interval estimator for r, it can be shown that

$$\frac{\sqrt{n}\,(\hat{r}(b) - r)}{\sigma(b)/\overline{\tau}} \xrightarrow[n \to \infty]{D} N(0,1),$$

where $\xrightarrow[n \to \infty]{D}$ denotes convergence in distribution and $\hat{\sigma}^2(b) \equiv \mathrm{var}(Y_j - r\tau_j - b\mathbf{X}_j)$. If we replace $\sigma(b)$ by an asymptotically consistent estimator s(b) then the same convergence applies and we can construct confidence intervals. However, we typically do not know the optimal values of b. The optimal value of b is β and is given by

$$\beta = \rho(Y - r\tau, \mathbf{X})\Sigma_{\mathbf{XX}}^{-1}, \tag{2.2.41}$$

where $\Sigma_{\mathbf{XX}}$ is the $Q \times Q$ covariance matrix of controls and $\rho(Y - r\tau, \mathbf{X})$ is the $1 \times Q$ row vector of covariances between $Y - r\tau$ and $\mathbf{X}$. The vector β minimizes the variance of $\hat{r}$. An estimate of β , $\hat{\beta}$ is obtained by using the appropriate sample covariances in (2.2.41). The variance of $\hat{r}$ can be estimated using

$$\hat{V}(\hat{r}) = s^2(\hat{\beta})/(\bar{\tau}\sqrt{n})^2,$$

where

$$s^2(\hat{\beta}) = \frac{1}{n-1}\sum_{j=1}^{n}\left[Y_j - \hat{r}\tau_j - \hat{\beta}(X_j - \bar{X})\right]^2.$$

An asymptotically correct $100(1-\alpha)\%$ confidence interval is given by

$$\hat{r} \pm \frac{Z_{\alpha/2}s(\hat{\beta})}{\sqrt{n}\,\bar{\tau}}$$

Lavenberg, Moeller and Sauer (1979) describe a specialized set of ratio-type controls for regenerative estimators.

2.2.3 Analysis Techniques for Nonnormal Responses

Lavenberg, Moeller and Welch (1982) present a method of producing confidence intervals based on the jackknife statistic, and they apply this method in a Monte Carlo study of a broad class of closed queueing networks.

Let $\bar{Y}_{(k)}(\hat{\beta})$ be the estimator computed (2.2.16) using the methodology of Lavenberg and Welch (1981) when the k^{th} observation has been deleted.

Compute the "pseudovalues"

$$J_k = K\,\bar{Y}(\hat{\beta}) - (K-1)\bar{Y}_{(k)}(\hat{\beta}), \quad 1 \leq k \leq K\ . \tag{2.2.42}$$

Now calculate the jackknife statistic

$$\bar{J}(\hat{\beta}) = \frac{1}{K} \sum_{k=1}^{K} J_k(\hat{\beta})\ , \tag{2.2.43}$$

and the sample variance

$$S_J^2(\hat{\beta}) = \frac{1}{K-1} \sum_{j=1}^{K} (J_j(\hat{\beta}) - \bar{J}(\hat{\beta}))^2 \tag{2.2.44}$$

An asymptotically valid confidence interval is given by

$$\bar{J}(\hat{\beta}) \pm\ t_{K-1}(1-\alpha) S_J(\hat{\beta})/\sqrt{K}\ . \tag{2.2.45}$$

We are referred to Arvensen (1969) for proof. These intervals hold under mild regularity conditions given in Miller (1974).

2.2.4 Experimental Results

Lavenberg, Moeller and Welch (1982) apply the control-variate confidence intervals (2.2.39) and (2.2.45) across a general class of closed queueing networks. They develop three types of controls, service time variables, flow variables, and work variables. The networks considered take the following form. Consider a finite set (say of size S) of interconnected service centers. These centers service D different types of customers. There are a total of N customers of all types. Assume

1. Markovian Routing so that the next station visited only depends on the current location.

2. The service times for the the j^{th} type of customer at the i^{th} service station are drawn independently from identical populations with finite mean and variance.

3. Service time sequences and sequences of centers visited are mutually independent.

The above networks form a general class of closed queueing networks. Since the only random components of this system are derived from the service time distributions and the multinomial routing distributions, functions of these variables can be used as internal controls.

The authors perform a rather extensive study across many different networks of the type described above. Three response variables were studied separately: the long run average waiting time (by customer type), the long run rate at which departures occur (by customer type), and the long run average response time (by customer type). The following are important

conclusions of their work:

1. Work variables exhibit the smallest minimum variance ratios. (The authors expected this since these variables contain both information on service time and flow).

2. The loss factor derived in Lavenberg and Welch (1978) appeared to adjust the minimum variance ratio correctly.

3. The actual coverage probability for nominal $100(1-\alpha)\%$ confidence intervals did not suffer with the application of controls.

4. The regression method produced substantially smaller confidence intervals than the jackknife method (with no appreciable degradation in coverage).

5. The forward selection procedure (Draper and Smith (1981)) was used to cope with the control variable selection problem.

Wilson and Pritsker (1984a,b) offer theoretical and experimental results on what they call "standardized" concomitant variables. Assume we are dealing with a Q-station queueing network. Define the input processes as $\{(U_j(k) : j \geq 1)\}$, $1 \leq k \leq Q$. Control variables will necessarily be functions of these inputs. Consider a control of the form

$$X_k(t) = (1/a(k,t)) \sum_{j=1}^{a(k,t)} (U_j(k) - \mu_k) , \tag{2.2.46}$$

where $a(k,t) \equiv$ number of service times that are started at station k during the time period [0,t]. Also μ_k is the known mean of the k^{th} control. Wilson

(1982) showed that controls of the type given by (52) have asymptotic mean and variance equal to zero. He states that this result also applies to the "work" variables given by Lavenberg, Moeller, and Welch (1982). As a consequence of this fact, the covariance matrix of the controls becomes asymptotically singular. The authors offer remedy in the form of standardized controls. Consider controls of the following form:

$$X_k(t) = (a(k,t))^{-1/2} \sum_{j=1}^{a(k,t)} (U_j(k) - \mu_k)/\sigma_k \ . \tag{2.2.47}$$

Here σ_k is the known standard deviation of the k^{th} control. The vector of standardized controls is shown to converge to a multivariate normal distribution with zero mean vector and identity covariance matrix, as the run length goes to infinity.

One may standardize the "work" variables given in Lavenberg, Moeller,and Welch (1982) by defining the controls as

$$X_k^*(t) = (\sqrt{f(k,t)}/\omega_k(f(t))) \sum_{j=1}^{f(k,t)} (U_j(k) - \mu_k)/\sigma_k \ , \tag{2.2.48}$$

where $\omega_k \equiv$ relative frequency with which a customer visits station k and $f(k,t) \equiv$ number of service times that are finished at station k during time period (0,t).

Wilson and Pritsker develop a theory of controlled replication analysis which is based on the asymptotic multivariate normality assumption and present selected simulation results. A more thorough experimental treatment

is given in Wilson and Pritsker (1984b). The variance reductions observed in both papers certainly offer compelling evidence that further research in these areas may prove extremely fruitful. Experiments were carried out for controlled replication and controlled regeneration analysis. The systems studied were a class of closed and mixed queues representing machine-repair systems. In the controlled replication experiments, variance reductions in the range from 20% to 90% were observed with confidence interval reductions ranging from 10% to 70%. After the effects of initialization bias were removed, no significant loss in coverage was observed. In the controlled regenerative experiments, variance reductions in the 30% to 90% range were observed with confidence interval reductions of 20% to 65%. Some coverage difficulties were noted (probably due to the inherent bias of the regenerative estimator) but degradation seemed to stay within about 10% of nominal.

2.3 Univariate Simulation Metamodel with Multiple Controls

In all the papers reviewed to this point, we have been working with a single underlying population and sampling from it. If we allow the population to vary, say over the design points of an experimental design, then we are working in the multipopulation domain. Nozari, Arnold and Pegden (1984) discuss the application of control variables for multipopulation experiments. These experiments typically involve some form of a general linear model. This model represents a simplification of the simulation and, as such, it is often called a metamodel. One object of multipopulation experiments is to find a metamodel which can closely predict a response across the domain of factor levels. Factor levels here correspond to to the design points mentioned earlier. Another object of such experiments is to identify as closely as

possible the values of the metamodel coefficients. Such information may tell us the relative sensitivity of a response to a particular factor. We also may glean whether or not certain factors should be included in the metamodel. This latter objective is the thrust of the research put forward by Nozari et al.

Let $\mathbf{Y} = (Y_1, \ldots, Y_K)'$ be a $K \times 1$ vector of independent observations. Each Y_i is obtained from an independent run of the model and each has common variance σ^2. Assume

$$\mathbf{Y} \sim N_K(Z\beta, \sigma^2 I_K),$$

where Z is a $K \times m$ known matrix of rank m, β is a $m \times 1$ vector of unknown coefficients and I_K is the $K \times K$ identity matrix. The factors or functions thereof are embodied in Z, when the factor levels are not random the matrix Z is commonly called the design matrix. Let X_i be a $Q \times 1$ vector of controls for the i^{th} observation of Y. Assume X_i has a known mean vector. Without loss of generality, we assume $E(X_i)=0$. Finally assume

$$\begin{pmatrix} Y_i \\ X_i \end{pmatrix} \sim N_{Q+1}\left(\begin{pmatrix} \mu_i \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & \Sigma_{YX} \\ \Sigma_{XY} & \Sigma_{XX} \end{pmatrix} \right) \tag{2.3.1}$$

where $E(Y_i) = \mu_i$ and Σ_{YX} and Σ_{XX} are covariance matrices of the response and the controls and the controls with themselves, respectively. Let Σ denote the $(Q+1) \times (Q+1)$ covariance matrix in (2.3.1). Assuming the metamodel has been correctly specified, Nozari et al. derive expressions for

Scheffe' type simultaneous confidence intervals for both the case when Σ is known and the case when Σ is unknown. In practice Σ is unknown and must be estimated. In this case, let A be a $(m-h)\times m$ matrix of rank m-h. Let

$$G = (Z \quad \mathbf{X}) \, , \mathbf{X} = (X_1, \, \ldots, \, X_K)',$$

where G is $K\times(m+Q)$ of rank $(m+Q)$. Now for simultaneous confidence intervals

$$\Pr\left\{ v'A\beta \in v'A\hat{\beta}_{CV} \pm \hat{\tau}\left[(m-h)F_{m-h,K-m-Q}(\alpha)v'A\left[G'G\right]^{-1}A'v\right]^{\frac{1}{2}} \right.$$

$$\left. \forall \; v \in \mathbb{R}^{m-h} \right\} = 1-\alpha,$$

where

$$\hat{\beta}_{CV} = (I_m \quad O)\,(G'G)^{-1}G'Y,$$

and

$$\hat{\tau}^2 = \frac{||\mathbf{Y} - G(G'G)^{-1}G'\mathbf{Y}||^2}{K-m+Q},$$

and $F(\alpha)$ is the alpha percentage point of a F distributed random variable also $(I_m \quad O)$ is $m\times(m+Q)$ with I_m the $m\times m$ identity matrix.

Nozari et al. define efficiency as

$$\mathrm{var}(\hat{\beta}) - \mathrm{var}(\hat{\beta}_{CV}) = \left[\sigma_\epsilon^2 - \tau^2 \frac{K-m-1}{K-m-Q-1}\right](Z'Z)^{-1},$$

where $\hat{\beta}$ are the estimated coefficients when controls are not used and $\tau^2 = \sigma_\epsilon^2(1 - \rho_{YX}^2)$. In similar fashion to Lavenberg et al. (1982), the loss factor here is

$$\frac{K-m-1}{K-m-Q-1}.$$

2.4 Multiresponse Simulation with Multiple Controls

Rubinstein and Marcus (1985) extend the development of Lavenberg, Moeller, and Welch (1982) to the estimation of a multivariate mean response using multiple controls. Extending to p responses, we see that the controlled estimator becomes

$$\mathbf{Y}(B) = \mathbf{Y} - B(\mathbf{X} - \mu_{\mathbf{X}}) , \tag{2.4.1}$$

where $\mathbf{Y}$ is a $p \times 1$ vector of responses, B is the $p \times q$ matrix of control coefficients, and $\mathbf{X}$ is a $q \times 1$ vector of controls with mean vector μ_x. Rubinstein and Marcus demonstrate that $\det(\mathrm{cov}(\mathbf{Y}(B)) = |\mathrm{cov}(\mathbf{Y}(B))|$, the generalized variance of $\mathbf{Y}(B)$, is minimized by

$$\beta = \Sigma_{YX}\Sigma_{XX}^{-1} , \tag{2.4.2}$$

where

$$\Sigma_{YX} = E((\mathbf{Y}-\mu_Y)(\mathbf{X}-\mu_X)') , \tag{2.4.3}$$

and

$$\Sigma_{XX} = E((\mathbf{X}-\mu_X)(\mathbf{X}-\mu_X)') . \tag{2.4.4}$$

The resulting minimum generalized variance is given by

$$|\mathrm{cov}\mathbf{Y}(\beta)| = |\Sigma_{YY} - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}| = |\Sigma_{YY}| \prod_{i=1}^{\nu}(1-\rho_i^2) , \tag{2.4.5}$$

where $\nu = \mathrm{rank}(\Sigma_{YX})$ and ρ_i^2, $i = 1, \ldots ,\nu$ are the canonical correlations between **Y** and **X** that satisfy $\rho_1 \geq \ldots \geq \rho_\nu$.

The authors define efficiency of control variables as

$$\delta_I^2 = \frac{|\mathrm{var}(\mathbf{Y}(\beta))|}{|\mathrm{var}(\mathbf{Y})|} = \frac{|\Sigma_{Y(B)}|}{|\Sigma_Y|} . \tag{2.4.6}$$

Porta Nova (1985) points out that the use of the term efficiency might not be warranted here, because one seeks an increase in an efficiency measure, whereas we seek to decrease δ_I^2 . In any case, δ_I^2 measures the relationship between **Y** and **X**. One can see by examination of (2.4.5) that the larger the canonical correlations, the greater the reduction in the generalized variance.

Now given

$$\mathbf{Z} = \begin{bmatrix} \mathbf{Y} \\ \mathbf{X} \end{bmatrix} \sim N_{p+Q}\left(\begin{bmatrix} \mu_Y \\ \mu_X \end{bmatrix}, \begin{bmatrix} \Sigma_{YY} & \Sigma_{YX} \\ \Sigma_{XY} & \Sigma_{XX} \end{bmatrix} \right), \tag{2.4.7}$$

we take a random sample of k observations

$$\mathbf{Z}_i = \begin{bmatrix} \mathbf{Y}_i \\ \mathbf{X}_i \end{bmatrix}, \quad i = 1, \ldots, K. \tag{2.4.8}$$

The authors consider two cases :

1. The matrix β is known.

2. The matrix β is unknown and must be estimated. Case (1) can be used to derive the multiresponse analogue η_2^* of the minimum variance ratio η_1^* (see equation (2.2.13)), while case (2) is used to derive the multiresponse analogue λ_2 of the loss factor λ_1 (see equation (2.2.10)). Both then are to be brought together to form what the authors call efficiency and Venkatraman and Wilson (1986) call the variance ratio (η_2).

Examination of (2.4.5) and (2.4.6) shows that the minimum variance ratio is given by

$$\delta_1^2 = \eta_2^* = \prod_{i=1}^{\nu} (1 - \rho_i^2) , \tag{2.4.9}$$

where ν and ρ are as in (2.4.5). Rubinstein and Marcus point out that η_2^*

can be measured by the ratio of the squared volume V_1^2 of the confidence ellipsoid formed about the estimate of μ_Y using control variables to the squared volume V_2^2 of the ellipsoid formed by direct simulation. Specifically, (2.4.7) implies that

$$K(\overline{\mathbf{Y}}-\mu_{\mathbf{Y}})'\Sigma_{\mathbf{YY}}^{-1}(\overline{\mathbf{Y}}-\mu_{\mathbf{Y}})\sim\chi_p^2, \tag{2.4.10}$$

where χ_p^2 is a chi-squared random variable with p degrees of freedom.

Hence we can form a $100(1-\alpha)\%$ confidence ellipsoid for $\mu_{\mathbf{Y}}$ from

$$Pr(K(\overline{\mathbf{Y}}-\mu_{\mathbf{Y}})'\Sigma_{\mathbf{YY}}^{-1}(\overline{\mathbf{Y}}-\mu_{\mathbf{Y}}) \leq \chi_{p,1-\alpha}^2) = 1-\alpha\ , \tag{2.4.11}$$

where

$$\overline{\mathbf{Y}} = \frac{1}{K}\sum_{j=1}^{K}\mathbf{Y}_j\ . \tag{2.4.12}$$

The volume of this ellipsoid is given by

$$V_1 = p^{-1}C(p)\,|\,\Sigma_{\mathbf{YY}}|^{1/2}(\chi_{p,1-\alpha}^2/K)^{p/2}\ , \tag{2.4.13}$$

where

$$C(p) = 2\pi^{p/2}/\Gamma(p/2).$$

We can also form an ellipsoid based on the controlled estimator from

$$\Pr(K(\overline{\mathbf{Y}}-\mu_Y)'\Sigma_{Y(\beta)}^{-1}(\overline{\mathbf{Y}}-\mu_Y) \leq \chi^2_{p,1-\alpha}) = 1-\alpha. \tag{2.4.14}$$

The volume of this ellipsoid is

$$V_2 = p^{-1}C(p)\,|\Sigma_{Y(\beta)}|^{1/2}(\chi^2_{p,1-\alpha})^{p/2}\ . \tag{2.4.15}$$

We note that the squared ratio of (2.4.13) to (2.4.15) produces

$$\delta_1^2 \equiv \eta_2' = \frac{V_1^2}{V_2^2} = \frac{|\Sigma_{Y(\beta)}|}{|\Sigma_{YY}|}, \tag{2.4.16}$$

which verifies (2.4.9).

In practice Σ_Z (the covariance matrix for Z given in (2.4.7)) is unknown and it must be estimated. Let S_Z denote the sample covariance matrix

$$S_Z = \begin{pmatrix} S_{YY} & S_{YX} \\ S_{XY} & S_{XX} \end{pmatrix} \tag{2.4.17}$$

where, for example,

$$S_{YX} = \frac{1}{K-1}\sum_{j=1}^{K}(\mathbf{Y}_j-\overline{\mathbf{Y}})(\mathbf{X}_j-\overline{\mathbf{X}})', \tag{2.4.18}$$

where $\overline{\mathbf{Y}}$ is given by (66) and $\overline{\mathbf{X}}$ is constructed analogously. Now we estimate β by $\hat{\beta}=S_{YX}S_{XX}^{-1}$ and form the K controlled responses as

$$\mathbf{Y}_j(\hat{\beta}) = \mathbf{Y}_j-\hat{\beta}(\mathbf{X}_j-\mu_X),\ \ 1 \leq j \leq K. \tag{2.4.19}$$

Under our assumption of multivariate normality an unbiased estimator of $\mu_{\mathbf{Y}}$ is

$$\overline{\mathbf{Y}}(\hat{\beta}) = \frac{1}{K}\sum_{j=1}^{K}\mathbf{Y}_j(\hat{\beta}) = \overline{\mathbf{Y}} - \hat{\beta}(\overline{\mathbf{X}} - \mu_{\mathbf{X}}) \ . \tag{2.4.20}$$

We can form a $100(1-\alpha)\%$ confidence interval from the relationship (Rao (1967))

$$\Pr\Big\{(\overline{\mathbf{Y}}(\hat{\beta}) - \mu_{\mathbf{Y}})'\hat{\Sigma}_{Y|X}^{-1}(\overline{\mathbf{Y}}(\hat{\beta}) - \mu_{\mathbf{Y}}) \leq$$

$$(\mathbf{d}'\mathbf{d})[(K-Q-1)p/(K-Q-p)]F_{p,K-Q-p}(1-\alpha)\Big\} = 1-\alpha,$$

where

$$\mathbf{d}' = \mathbf{1}'_K/K - (\overline{\mathbf{X}} - \mu_{\mathbf{X}})'(\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}',$$

where G is defined in (2.4.25) and $\mathbf{1}_K$ is a K dimensional column vector of ones. Also we have

$$\hat{\Sigma}_{\mathbf{Y}|\mathbf{X}} = \frac{K-1}{K-Q-1}(S_{\mathbf{YY}} - S_{\mathbf{YX}}S_{\mathbf{XX}}^{-1}S_{\mathbf{XY}}).$$

Rubinstein and Marcus define efficiency of control variables as

$$\epsilon_1^2 = \frac{E\Big\{|\hat{\Sigma}_{\mathbf{Y}|\mathbf{X}}|(\mathbf{d}'\mathbf{d})^p\Big\}}{E\Big\{(1/K)^p\,|S_{\mathbf{YY}}|\Big\}},$$

this is the ratio of the expected generalized sample variance of $\overline{\mathbf{Y}}(\hat{\beta})$ to the expected generalized variance of $\overline{\mathbf{Y}}$. They prove that

$$\epsilon_1^2 = C_1(K,Q,p)C_2(K,Q,p)\prod_{i=1}^{\nu}(1-\rho_i^2)$$

where $\nu = rank(\Sigma_{\mathbf{YX}})$ and

$$C_1(K,Q,p) = \prod_{i=1}^{p}\left[(K-Q-i)(K-1)/(K-Q-1)(K-i)\right]$$

and

$$C_2(K,Q,p) = 1+\sum_{j=1}^{p}\binom{p}{j}\frac{Q(Q+2)\ \dots\ (Q+2(j-1))}{(K-Q-2)\ \dots\ (K-Q-2j)}$$

we see immediately that the loss factor for this measure of efficiency is

$$C_1(K,Q,p)\ C_2(K,Q,p).$$

Venkatraman and Wilson derive a more natural extension of the loss factor of Lavenberg, Moeller, and Welch (1982) by simply calculating the "efficiency" or variance ratio as

$$\eta_2 \equiv \eta_2^* \lambda_2\ , \tag{2.4.21}$$

where η_2 is the variance ratio given by

$$\eta_2 = \frac{\mathrm{var}(\overline{\mathbf{Y}}(\hat{\beta}))}{\mathrm{var}(\overline{\mathbf{Y}})} = \left[\frac{K-2}{K-Q-2}\right]^p \prod_{i=1}^{\nu}(1-\rho_i^2), \tag{2.4.22}$$

η_2^* is the minimum variance ratio given by

$$\eta_2^* \equiv \frac{\mathrm{var}(\overline{\mathbf{Y}}(\beta))}{\mathrm{var}(\overline{\mathbf{Y}})} = \prod_{i=1}^{\nu}(1-\rho_i^2), \tag{2.4.23}$$

and λ_2 is the loss factor given by

$$\lambda_2 \equiv \frac{\mathrm{var}(\overline{\mathbf{Y}}(\hat{\beta}))}{\mathrm{var}(\overline{\mathbf{Y}}(\beta))} = \left[\frac{K-2}{K-Q-2}\right]^p. \tag{2.4.24}$$

To calculate λ_2 we need an expression for $\mathrm{var}(\overline{\mathbf{Y}}(\hat{B}))$. One way to calculate this is to write $\overline{\mathbf{Y}}(\hat{\beta})$ as a linear combination of the uncontrolled responses and then calculate the covariance matrix of this vector. This procedure is an extended variant of the procedure used in Lavenberg, Moeller, and Welch (1982).

Let $\mathbf{G}$ be defined as

$$\mathbf{G} = \begin{bmatrix} (\mathbf{X}_1-\overline{\mathbf{X}})' \\ \cdot \\ \cdot \\ \cdot \\ (\mathbf{X}_K-\overline{\mathbf{X}})' \end{bmatrix} \tag{2.4.25}$$

where $\mathbf{X}_i$, $1 \leq i \leq K$ is a $Q\times 1$ column vector of observed controls and $\bar{\mathbf{X}}$ is the $Q\times 1$ vector of the sample control means.

Analogous to equations (2.2.18), (2.2.19), and (2.2.20) we can write

$$\bar{\mathbf{Y}}(\hat{\beta}) = \bar{\mathbf{Y}} - \hat{B}(\bar{\mathbf{X}}-\mu_{\mathbf{X}}) = (\mathbf{Y}_1,\ldots,\mathbf{Y}_K)\mathbf{H} \ , \tag{2.4.26}$$

where $(\mathbf{Y}_1,\ldots,\mathbf{Y}_K)$ is the matrix of observed uncontrolled responses and $\mathbf{H}$ is defined as

$$\mathbf{H} \equiv \frac{1}{K}\mathbf{1}_K - \mathbf{G}(\mathbf{G}'\mathbf{G})^{-1}(\mathbf{X}-\mu_{\mathbf{X}}) \tag{2.4.27}$$

where $\mathbf{1}_K$ is a K dimensional column vector of ones. The authors show that $\bar{\mathbf{Y}}(\hat{\beta})$ is an unbiased estimator of $\mu_{\mathbf{Y}}$.

$\mathrm{var}(\bar{\mathbf{Y}}(\hat{\beta}))$ is calculated via the law of total probability for expectations. First, we calculate the conditional covariance of $\bar{\mathbf{Y}}(\hat{\beta})$ given the observed controls, then we take the expectation of this quantity across the controls. This double expectation is the correct quantity since the unconditional and conditional expectations of $\bar{\mathbf{Y}}(\hat{\beta})$ are the same.

After both expectations are taken we get

$$\mathrm{var}(\bar{\mathbf{Y}}(\hat{\beta})) = \frac{K-2}{K(K-Q-2)}\Sigma_{\mathbf{Y}|\mathbf{X}} \ , \tag{2.4.28}$$

where

$$\Sigma_{\mathbf{Y}|\mathbf{X}} = \Sigma_{\mathbf{YY}} - \Sigma_{\mathbf{YX}}\Sigma_{\mathbf{XX}}^{-1}\Sigma_{\mathbf{XY}} \ . \tag{2.4.29}$$

Now

$$\operatorname{cov}(\overline{\mathbf{Y}}(\hat{\gamma})) = \frac{1}{K} \Sigma_{\mathbf{Y}|\mathbf{X}} \; . \tag{2.4.30}$$

so the loss factor λ_2 is

$$\lambda_2 \equiv \frac{|\operatorname{cov}\overline{\mathbf{Y}}(\hat{\beta})|}{|\operatorname{cov}\overline{\mathbf{Y}}(\hat{\gamma})|} = \left(\frac{K-2}{K-Q-2} \right)^p \tag{2.4.31}$$

η_2^* is given by (2.4.23); hence the generalized variance ratio η_2 is given by

$$\eta_2 = \left(\frac{K-2}{K-Q-2} \right)^p \prod_{i=1}^{\nu} (1-\rho_i^2) \; . \tag{2.4.32}$$

Obviously we require $K-Q-2 > 0$.

Venkatraman and Wilson provide guidance for limiting the number of controls. They state that if Λ is a user specified upper limit on the loss λ_2, then at most

$$Q^* = \left[(K-2)(1 - \Lambda^{-1/p}) \right]$$

controls should be applied.

2.5 Multiresponse Simulation Metamodel with Multiple Controls

Porta Nova (1985) extends the development of Nozari, Arnold and Pegden (1984) to the case of multiresponse metamodels with multiple controls.

The model employed is as follows

$$\mathbf{Y} = Z\beta + \mathbf{X}\Delta + R$$

where $\mathbf{Y}$ is a $K\times p$ matrix consisting of K p-dimensional observations, Z is a $K\times m$ design matrix, β is a $m\times p$ matrix of unknown parameters, $\mathbf{X}$ is a $K\times Q$ matrix consisting of K Q-dimensional vectors of controls, Δ is a $Q\times p$ matrix of unknown control coefficients and R is a $K\times p$ matrix of residuals.

Porta Nova provides point estimators for β and Δ

$$\hat{\beta} = (Z'Z)^{-1}\left[I - X(X'\mathrm{P}X)^{-1}X'\mathrm{P}\right]Y,$$

and

$$\hat{\Delta} = (X'\mathrm{P}X)^{-1}X'\mathrm{P}Y$$

where

$$\mathrm{P} = I - Z(Z'Z)^{-1}Z'.$$

A $100(1-\alpha)\%$ confidence ellipsoid for vec β (vec β is a column vector

obtained by concatenating the columns of β) is derived from

$$\text{vec } (\hat{\beta} - \beta)'(\hat{\Sigma}_{Y|X} \otimes BB')^{-1} \text{ vec } (\hat{\beta} - \beta) \quad |X \equiv T^2_{mp}(K-m-Q)$$

where $T^2_{mp}(K-m-Q)$ is Hotelling's T^2 with $K-m-Q$ degrees of freedom, here

$$B = (Z'Z)^{-1}Z'\left[I - X(X'PX)^{-1}X'P\right]$$

and

$$\hat{\Sigma}_{Y|X} = R'R/(K-m-Q)$$

where

$$R = Y - \left[Z\hat{\beta} + X\hat{\Delta}\right]$$

and $\otimes$ is the Kronecker product. The Kronecker product of the $m \times n$ matrix A with the $p \times q$ matrix B is the $mp \times nq$ matrix

$$A \otimes B = \begin{bmatrix} A_{11}B & \cdots & A_{1n}B \\ & \vdots & \vdots \\ A_{m1}B & \cdots & A_{mn}B \end{bmatrix}.$$

A $100(1-\alpha)\%$ confidence ellipsoid about vec β is formed from

$$\Pr\left\{\frac{K-m-Q-mp+1}{mp(K-m-Q)}T^2_{mp}(K-m-Q) \le F_{mp,K-m-Q-mp+1}(1-\alpha)\right\} = 1-\alpha$$

Porta Nova generalizes the minimum variance ratio as

$$\eta_3^* = \left[\prod_{i=1}^{\nu}(1-\rho_i^2)\right]^m,$$

he also provides a loss factor of

$$\lambda_3 = \left[\frac{K-m-1}{K-m-Q-1}\right]^p,$$

where $\nu = \text{rank}\Sigma_{\mathbf{YX}}$ and ρ_i are the canonical correlations between **Y** and **X**.

2.6 Selection of Regression Models

Typically a practitioner is confronted with multiple candidates for controls. It is possible (in view of the loss factor) that if he elected to use all the candidates, he might actually induce variance into his estimator. Therefore, a control variate selection procedure that finds the "best" subset of controls is desirable. We have seen that the principal technique employed to exploit control variates is linear regression. In our application, control variates are used as predictors in a linear regression on some response of interest. The statistical literature is replete with papers that deal with the selection of predictor variables in the regression context. However, relatively little has been written on the selection of controls. In this section we review control variate selection techniques as well as the more general literature on

the selection of predictor variables in linear regression.

2.6.1 Review of Control Variate Selection Techniques

Lavenberg, Moeller and Welch (1981) applied a restricted variant of "forward selection" regression (Draper and Smith (1981), Chap. 6) to selecting a "best" set of controls. Nozari, Arnold and Pegden (1984) developed variants of the "all regressions" and "forward selection" procedures. These variants were tailor-made for the selection of controls in the situation where the objective was the estimation of a univariate simulation metamodel with multiple controls. Porta Nova (1985) and Venkatraman and Wilson (1986) offer advice on the number of controls to be used but do not discuss how to select these controls.

2.6.2 Review of Variable Selection Techniques

There are several good survey articles written on the variable (predictor) selection problem. Draper and Smith (1981), Thompson (1978) and Hocking (1976) provide detailed surveys of the variable selection problem itself, while Hocking (1983) offers a succinct overview of the general topic of linear regression. Siotani, Hayakawa and Fujikoshi (1985) provide some multivariate extensions of selected methods.

To organize the discussion we will classify the selection techniques according to model and objective. Model will refer to either multiple linear regression (univariate response) or multivariate linear regression (multivariate response). The objective of a model will be classified as one of the following: prediction, description or control. Aitkin (1977) makes the distinction

between description and prediction. We draw attention to these objectives only to underscore our objective, control.

2.6.2.1 Multiple Linear Regression Model

Basic Model. The multiple linear regression model takes on the following form

$$y_i = X_i\beta + \epsilon_i \ , 1 \leq i \leq K$$

where y_i is the i^{th} independent observed response, X_i is the i^{th} observed $1{\times}(Q+1)$ row vector of predictors, β is a $(Q+1){\times}1$ column vector of unknown parameters (with a 1 in the first column) and ϵ_i is the i^{th} residual. We assume that $E(\epsilon_i) = 0$ and $\text{var}(\epsilon_i) = \sigma^2$, $1 \leq i \leq K$. We note that

$$E(y) = X\beta$$

and

$$\text{var}(y) = \sigma^2,$$

if we can take X as fixed. This is never the case when we employ control variates. We must be careful to distinguish between the case when X is fixed and the case when X is a random matrix. Following Thompson (1978), we impose the following condition for X random. Assume y, $X = x_1, \ldots, x_Q$ are jointly distributed as a (Q+1)-dimensional normal distribution with unknown mean vector and covariance matrix. The conditional expectation of y given

$X = x$ is

$$E(y \mid X = x) = X\beta$$

$$\mathrm{var}(y \mid X = x) = \sigma^2$$

where β and σ^2 can be expressed in terms of the mean vector and covariance matrix (see Anderson (1984), Chap. 2). Now conditionally on $X = x$, y is normally distributed with mean and variance as given above. Finally, conditioned on $X = x$, the model is

$$y = X\beta + \epsilon, \tag{2.6.1}$$

where X is $1\times(Q+1)$, β is $(Q+1)\times 1$ and $E(\epsilon) = 0$ with $\mathrm{var}(\epsilon) = \sigma^2$. We will see that certain expectations can be computed over X in the case of random predictors that lead to criteria for variable selection.

It turns out that the estimates used for β and σ^2 are the same for either model, however; the distributions of the estimators differ. If we array all the K observations we have

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon$$

where $\mathbf{Y}$ is $K\times 1$, $\mathbf{X}$ is $K\times(Q+1)$, β is $(Q+1)\times 1$ and ϵ is $K\times 1$. The least squares solution for β is

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

An unbiased estimator of σ^2 for both methods is given as

$$\hat{\sigma}^2 = \frac{1}{K-Q-1}(\mathbf{Y} - \mathbf{X}\hat{\beta})'(\mathbf{Y} - \mathbf{X}\hat{\beta})$$

Selection of Variables

Given we have Q candidates for predictors, we wish to select a subset of predictors that is in some sense "best". Using the notation of Aitkin (1974), we assume that the first p variables are selected and the last $Q-p$ are eliminated. The subset model is

$$\mathbf{Y} = \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2 + \epsilon,$$

where

$$\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2),$$

and

$$\beta = (\beta_1, \beta_2).$$

Here β_0 is included in $\mathbf{X}_1$ so $\mathbf{X}_1$ is $K \times (p+1)$. If the last Q-p variables are not included in the model we take $\beta_2 = 0$.

The methods we survey are typically of the following form. A criterion is chosen that reflects the intended use of the regression model. Subsets are evaluated according to this criterion and the "best" subset is chosen as the solution. Most taxonomies (Draper and Smith (1981), Thompson (1978), Hocking (1976)) are dominated by three approaches: 1) All possible regressions, 2) best k subset regressions and 3) sequential procedures.

The all regressions approach entails a complete enumeration of all $2^Q - 1$ combinations of predictors, each evaluated according to a criterion. Once the criterion has been calculated for all subsets, the information is arrayed and a subjective choice is made based on the criterion values and, perhaps, auxiliary information. Clearly, this is a computation-intensive method. This method was not a tractable procedure until the advent of modern high-speed computers.

Best k subset regressions and sequential procedures have been developed in an effort to avoid the examination of all possible regression subsets. Draper and Smith (1981) cite the branch and bound algorithm given by Furnival and Wilson (1974) that computes only a small fraction of the possible regressions and yields the "best k " subsets (k is user specified). Sequential procedures are more economical than the all regressions approach in that they try to find the "best" regression of a certain number of variables. There is no guarantee that these methods will find the "best regression", in fact, different sequential procedures often yield different solutions.

The all possible regressions and best k subsets procedures use a specified selection criterion. There is nothing from preventing us from applying these criteria in a sequential fashion. However, we will only discuss sequential

procedures that are based on a sort of partial F-test. Our discussion of sequential procedures will concentrate on 1) Forward Selection, 2) Backwards Elimination and 3) Stepwise Regression. First we discuss the following five popular selection criteria: 1) the S_p criterion, 2) Akaike's Information Criterion (AIC), 3) R_p^2, 4) R_a^2 and 5) Mallow's C_p criterion. Next we discuss the three sequential procedures. We will end our discussion of variable selection techniques for the multiple linear regression model with a summary of some other methods: 1) Ridge regression, 2) Principal Components Regression, 3) Latent Root Regression, 4) Press and 5) Inferential techniques due to Aitken and McCabe.

Selection Criteria.

The S_p Criterion. The first selection criterion we discuss is the S_p criterion. Thompson (1978) recommends this selection criterion as preferable if the predictor variables and response can be taken to represent a (Q+1)-dimensional normal distribution. This criterion seeks to minimize the expected mean square error of prediction. Following Thompson, mean square error of prediction is given by

$$MSEP = E_y(y - \hat{y}_p)^2,$$

where y, $\hat{y}_p$ are respectively the observed and predicted values of the response that correspond to some subset of size p ($p \leq Q$) of predictors. It can be shown that

$$E_y(y-\hat{y}_p)^2 = \frac{\sigma_p^2}{K}(1+K+T^2),$$

where σ_p^2 is the residual variance of the p-variable equation and T^2 is a Hotelling's T^2 statistic that is a function of the regression sample and the p predictors. If we take the expectation of MSEP across all regression samples and predictors sets for the p variables we get

$$\mathcal{E}_p = \frac{\sigma_p^2}{K}\left[1+K+\frac{p(K+1)}{K-p-2}\right].$$

This is estimated by

$$E_p = \frac{SSE_p}{K(K-p)}\left[1+K+\frac{p(K+1)}{K-p-2}\right],$$

where SSE_p is the sum of squares due to error from the p variable regression. After some algebra and recognition that K is fixed for a particular experiment, E_p is simplified to

$$S_p \equiv E_p = \frac{SSE_p}{(K-p)(K-p-2)}.$$

Lindley (1967) offers a Bayesian version of this criterion.

Akaike's Information Criterion (AIC). Akaike (1973) proposed a criterion of the following form

$$AIC_q = -2\log(\ f\,(y;x,\hat{\theta})\ ) + 2q$$

where $f\,(y;x,\hat{\theta})$ is the p.d.f. (likelihood) function of y evaluated at $\hat{\theta}$, $\hat{\theta}$ are the maximum likelihood estimates of the q unknown parameters of the subset model. Akaike derived this criterion from information theoretic considerations. This criterion does not enjoy widespread acceptance as a selection criterion. Schwarz (1978) offers a Bayesian version of AIC.

The Coefficient of Multiple Determination R_p^2. The coefficient of multiple determination R_p^2 is defined (for a subset of size p) as

$$R_p^2 = 1 - \frac{SSE_p}{SSTO}$$

where SSTO is the total sum of squares for y. The quantity SSTO is constant for all possible regressions and SSE_p is monotonically decreasing in p (a useful fact exploited by Furnival and Wilson (1974)), hence, we do not seek, necessarily, to find the maximum R_p^2. Here we are looking for the point where adding additional predictors is not worthwhile because of a small relative change in R_p^2. This method is clearly subjective.

The Adjusted Coefficient of Determination R_a^2. The adjusted coefficient of determination R_a^2 is defined as

$$R_a^2 = 1 - \left(\frac{K-1}{K-p} \right) \frac{SSE_p}{SSTO}.$$

This criterion takes into account the subset size p and penalizes candidate predictors whose addition does not adequately reduce SSE_p. Neter, Wasserman and Kutner (1983) suggest graphical procedures for both R_p^2 and R_a^2.

Mallows' C_p Criterion. Mallows (1973) suggests a criterion based on minimizing MSEP in the case $X = X_1, \ldots, X_Q$ are fixed. Here we will assume that the "true" model contains all Q predictors. We seek to find a subset model (although biased due to misspecification, see Hocking (1976)) that provides a similar MSEP to the Q variable model and is nearly unbiased. It can be shown that

$$\text{MSEP}(\hat{y}_i) = \text{bias}(\hat{y}_i)^2 + \text{var}(\hat{y}_i).$$

If we total the mean square error for all K fitted values and divide by σ^2, the true error variance, we get

$$\Gamma_p = \frac{1}{\sigma^2}\left[\sum_{i=1}^{K}\left[E(\hat{y}_i)-E(y_i)\right]^2+\sum_{i=1}^{K}\text{var}(\hat{y}_i)\right],$$

a "standardized" total square error as a criterion. It can be shown that a good estimator of Γ_p is C_p where

$$C_p = \frac{SSE_p}{\hat{\sigma}^2}-(K-2p),$$

and

$$s^2 = \frac{SSE_Q}{K-Q}.$$

When there is no bias in the Q predictor model, $E(C_p) \simeq p$. If we plot C_p vs. p, models with substantial bias will tend to fall significantly above the line $C_p = p$. The strategy is to look for subsets with low bias and small C_p. Thompson shows C_p is closely related to R_p^2 and R_a^2. She points out that the C_p procedure tends to select a larger set of variables than R_p^2.

Sequential Procedures

In this section we will discuss three sequential variable selection techniques: 1) Forward Selection, 2) Backwards Selection and 3) Stepwise Regression. All these techniques are based on the "extra sum of squares" principle as related through the following theorem given in Thompson (1978).

Theorem. Given the linear model

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon,$$

where $\mathbf{Y}$ is $K\times 1$, $\mathbf{X}$ is $K\times Q$, β is $Q\times 1$ with ϵ a $K\times 1$ vector or residuals such that $E(\epsilon) = \mathbf{0}$ and $\mathrm{var}(\epsilon) = \sigma^2 I_K$. Then $\beta_{p+1} = \ldots = \beta_Q = 0$ implies that

$$F = \frac{SS_{(Q-p)}/Q-p}{MSE_Q} \sim F_{Q-p,K-Q}$$

where $SS_{Q-p} = SSR_Q - SSR_p$ and SSR_Q is the sum of squares due to

regression from all Q predictors.

We note that SS_{Q-p} is the extra sum of squares due to the Q-p extra variables in the model. The strategy of all three procedures in this section will involve successive partial F-tests that test the contribution to total sum of squares. These tests decide whether new variables enter or old variables leave the model.

Forward Selection. In the forward selection procedure we start with no variables in the model. The algorithm proceeds as follows:

1. Compute the sample partial correlation coefficient (w.r.t y) for all the predictors not in the model.

2. Choose as the entering candidate that predictor with the highest partial correlation with the response.

3. Perform a F-test based on a model of order equal to the current number of predictors in the model plus one for the entering variable.

4. If the F statistic is significant continue with step 1; otherwise terminate the procedure.

Backwards Selection. In the backward selection procedure we start with all Q predictors in the model. The algorithm proceeds as follows:

1. Treating each predictor as the last to enter, compute

$$F_i = \frac{SS_i}{MSE_Q}$$

for each i^{th} variable not yet deleted.

2. Choose the minimum F_i as the candidate to leave the model.

3. Compare the candidate's partial F to a critical F.

4. If the partial F is nonsignificant we delete the candidate from the model and continue with step 1; otherwise we terminate the procedure.

Stepwise Regression. In the stepwise regression technique we extend the forward selection technique to allow for the deletion of variables at each step. The algorithm proceeds as follows:

1. Proceed with steps 1-4 of the Forward Selection technique.

2. If a variable is included at step 4 of the Forward Selection technique then calculate partial F statistics for all the variables currently in the model.

3. If a partial F is below the critical value, delete this variable; go to step 1 in either case.

As we mentioned earlier there is no guarantee that these methods will arrive at the same (or "best" for that matter) solution. There is some evidence that backwards elimination is superior to forward selection (Mantel (1970)). Draper and Smith (1981) prefer the stepwise procedure.

Other Techniques

Ridge Regression. The technique of ridge regression was developed to counter the effects of multicollinearity. Multicollinearity can arise when some

subset of the predic r variables are highly correlated. In the linear model given in (2.6.1), these high correlations can cause the matrix $\mathbf{X'X}$ to be nearly nonsingular. Rather than use the standard least squares estimators for β, we introduce the biased "ridge" estimator due to Hoerl and Kennard (1970):

$$\hat{\beta}_R = (\mathbf{X'X} + cI_K)^{-1}\mathbf{X'Y} \text{ , } 0 \le c < \infty, \tag{2.6.2}$$

where c is an arbitrary constant used to perturb the diagonal of the $\mathbf{X'X}$ matrix and thereby, hopefully, eliminate the "ill-conditioning" of the $\mathbf{X}$ matrix.

In practice a graphical aid called a ridge trace is employed to help find a good value of c. The ridge trace is a graph of the estimated coefficients (using (2.6.2)) vs. values of c (typically $0 \le c \le 1$). The ridge trace is examined for a value of c where the estimated coefficients stabilize. This device can also be used as a variable selection tool by identifying those predictors with 1) unstable ridge traces and 2) coefficients close to zero. Draper and Smith (1981) are careful to point out that this technique is not usually used for variable selection. Thompson (1978) objects to to this method on two grounds: 1) the method is arbitrary in that it lacks a specific criterion and has no stopping rule and 2) the relative magnitudes of the predictor variables are ignored.

Principal Components Regression. In principal components regression the approach is to break the rows of the $\mathbf{X}$ matrix into its principal components (we discuss the principal components technique in more detail in a later section) and retain only those components that explain the greatest portion of the "variance" in $\mathbf{X}$. The technique can help to remove

the problem of multicollinearity (by reduction of the dimension of **X**) but suffers from forcing the investigator to work with predictors that are linear compounds of the original predictors. These linear compounds may be difficult to interpret.

Latent Root Regression. This method represents an attempt to improve upon the principal components method. In latent root regression we augment the "correlation" matrix of **X** (used to extract the components in principal components regression) with **Y**. Now we extract the latent roots (eigenvalues) and corresponding latent vectors (eigenvectors). Next, we array the eigenvalues and eigenvectors in tabular form. We look for pairs of eigenvalues and that coefficient of the associated eigenvector that corresponds to **Y** that are small. Webster (1974) provides guidelines for smallness. If the smallness criterion is met, the eigenvector(s) in question is (are) candidates for deletion from the model. Next, we estimate the parameters for the subset model, back-transform into the original variables and compare to least square estimates. It is important to remember that the new estimates are biased. We can perform candidate examination in a method similar to backwards elimination (details in Webster). Draper and Smith do not recommend this method to the majority of practitioners due to its inherent bias and complexity.

PRESS. Prediction Sum of Squares (PRESS) is a hybrid method proposed by Allen (1971). It is hybrid in the sense that it combines all possible regressions and residual error into a psuedo-jackknife procedure. For each subset of size p, we delete one observation at a time and use the remaining observations to predict the deleted response. At each deletion we compute the difference between the observed response and the prediction.

This quantity is called the predictive discrepancy. Once through the data, we sum the squares of the predictive discrepancies and proceed to the next size of subset. When all the subsets have been exhausted, we array the information and choose a model with a low sum of squares and subset size. Draper and Smith (1981) point out that, although the method helps detect influential data points, there is no set stopping rule and the method is computation-intensive.

Inferential Methods. Aitkin (1974) points out that application of sequential procedures (forward selection, backward elimination, and stepwise regression) suffers from a serious drawback. The Type I family error rate for the sequences of dependent F-tests is unknown. Aitkin offers a simultaneous test procedure that controls the family error rate in the subset selection problem. If a partition of $\mathbf{X} = (\mathbf{X}_1\ \mathbf{X}_2)$ ($\mathbf{X}$ is $K \times p_1$, $\mathbf{X}_2$ is $K \times p_2$ and $p_1 + p_2 = Q+1$) is prespecified then it is well known that if

$$y = \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2,$$

with the same assumptions as in section 2.6.2.1, then a likelihood ratio test for $\beta_2 = 0$ is based on

$$F = \frac{(R_{\mathbf{X}}^2 - R_{\mathbf{X}_1}^2)/p_2}{(1-R_{\mathbf{X}}^2)/K-Q-1},$$

where $R_{\mathbf{X}_1}^2$ is the squared multiple correlation of y with the predictors $\mathbf{X}_1$. F has a noncentral F distribution and a test can be constructed for $\beta_2 = 0$ by finding the appropriate critical point of the null distribution. Aitken finds a

simultaneous test for all partitions $\mathbf{X}_1$ by noting

$$U(\mathbf{X}_1) = p_2 F = \frac{(R_{\mathrm{X}} - R_{\mathrm{X}_1})^2}{(1 - R_{\mathrm{X}}^2)/K-Q-1)} \tag{2.6.3}$$

and defining

$$U = \max_{\mathrm{X}_1}\left\{U(\mathbf{X}_1)\right\}.$$

We have the simultaneous test for all partitions if we do not reject a particular partition when

$$U(\mathbf{X}_1) < C^{\alpha}_{K,Q},$$

where $C^{\alpha}_{K,Q}$ is the $100\alpha\%$ point of the null distribution of U. Examination of (2.6.3) shows the maximum of $U(\mathbf{X}_1)$ occurs when $\mathbf{X}_1$ consists only of the first column of $\mathbf{X}$, that is, the column of ones. In this case

$$U = \frac{R_{\mathrm{X}}^2}{(1 - R_{\mathrm{X}}^2)/(K-Q-1)}.$$

U/Q is noncentral F and the simultaneous test that does not reject the hypothesis $\beta_2 = 0$ for an arbitrary partition if

$$\frac{R_{\mathrm{X}}^2 - R_{\mathrm{X}_1}^2}{(1 - R_{\mathrm{X}}^2)/K-Q-1} < QF_{Q,K-Q-1}(\alpha).$$

Aitkin calls subsets not rejected, R^2 - adequate.

In the case where the regression equation is to be used for description rather than prediction, he offers simultaneous testing procedures for the cases of **X** fixed and random. He calls subset models, that are not rejected by this procedure, MSPE (Mean Squared Prediction Error)-adequate. Let E_0 be defined as the MSPE for the full model when **X** is fixed. Let E_1 be defined as the MSPE for the full model when **X** and y are jointly distributed as a multivariate normal distribution. The simultaneous procedures test the hypothesis H_0: $E_i - E_i^* = 0$ vs. H_1: $E_i - E_i^* < 0$ (i=0,1,2 , Aiktin includes the case i=1 where **X** is taken to be distributed uniformly over x_i) where E_i^* is MSPE for a subset model. The approach is to classify a subset as MSPE-adequate if the MSPE for the subset is not significantly larger than the MSPE of the complete model. The test statistic for H_0: $E_0 - E_0^* = 0$ is

$$F_0' = \frac{\hat{\beta}'_2 \mathbf{X}_{2.1} \mathbf{X}'_{2.1} \hat{\beta}_2}{\hat{\sigma}^2 \mathbf{X}'_{2.1} S_{22.1}^{-1} \mathbf{X}_{2.1}},$$

where $\hat{\beta}_2$ are the least squares estimates (in the full model) of β_2, $\mathbf{X}_{2.1} = \mathbf{X}_2 - \bar{\mathbf{X}}_2 - S_{21} S_{11}^{-1} (\mathbf{X}_1 - \bar{\mathbf{X}}_1)$, S_{12} is the matrix of sample covariances between elements of $\mathbf{X}_1$ and $\mathbf{X}_2$, $S_{22.1} = S_{22} - S_{21} S_{11}^{-1} S_{12}$ and $\hat{\sigma}^2$ is an estimate of the residual error. The simultaneous test which does not reject the hypothesis H_0 for an arbitrary partition if

$$F_0' < Q F_{Q,K-Q-1}^{1}(\alpha),$$

where $F_{Q,K-Q-1}^{1}(\alpha)$ is the $100\alpha\%$ point of a noncentral F with Q and K-Q-1

degrees of freedom and noncentrality parameter of 1.

The test statistic for H_0: $E_2 - E_2' = 0$ is

$$R^2_{X_{2.1}} = \frac{(R^2_X - R^2_{X_1})}{1 - R^2_{X_1}}.$$

the simultaneous test that does not reject the hypothesis H_0 for an arbitrary partition if

$$R^2_{X_{2.1}} \leq h^{(p_2/(K-2))}_{p_2, K-p_1-1}(\alpha)$$

where $h^{(p_2/(K-2))}_{p_2, K-p_1-1}$ is the $100\alpha\%$ point of the distribution of the squared multiple correlation coefficient based on p_2 and $K-p_1-1$ degrees of freedom when the population squared multiple correlation is $p_2/(K-2)$.

McCabe (1978) proposes a framework for variable selection called α-acceptability. Basically, a subset is considered α-acceptable if the parameter estimates for the reduced model fall into the $100(1-\alpha)\%$ confidence region formed by the full model. We look for subset models that are "close" to the full model. Given the linear model of (2.6.1) it is well known that a $100(1-\alpha)\%$ confidence region for β is given by (see Johnson and Wichern (1982), pg 304)

$$S_\alpha = \left\{ b : D(b) < s^2(Q+1)F_{Q+1, K-Q-1}(1-\alpha) \right\},$$

where $D(b) = (\hat{\beta}_{ls} - b)'(X'X)(\hat{\beta}_{ls} - b)$. Here $\hat{\beta}_{ls}$ are the least squares estimates from the full model and s^2 is the estimate of residual variance ($s^2 = SSE(\hat{\beta}_{ls})$). We note D(b) is the squared distance of any estimator b from the least squares estimator $\hat{\beta}_{ls}$ in the units provided by $X'X$. Therefore values of b for which

$$D(b) < d_\alpha,$$

(where $d_\alpha = s^2(Q+1)F_{Q+1,K-Q-1}(1-\alpha)$) are α - acceptable. The α-acceptable subsets form a collection. If one of the subset models is the true model, McCabe provides a selection rule that guarantees that the probability that the correct model is included in the collection is greater than $1-\alpha$. The rule is to include all subsets for which

$$\frac{SSE(\hat{\beta}_s)}{SSE(\hat{\beta}_{ls})} \leq 1 + \frac{Q+1}{K-Q-1} F_{Q+1,K-Q-1}(1-\alpha),$$

where $\hat{\beta}_s$ are the coefficients in the subset model. This rule is derived from 1) the fact that $D(\hat{\beta}_s) = SSE(\hat{\beta}_s) - SSE(\hat{\beta}_{ls})$ and 2) for any $\beta = \beta_s$ (the true model is the subset model), $\Pr_{\beta}(\hat{\beta} \in S_\alpha) \geq 1 - \alpha$. McCabe notes that Aitkin's R^2-adequacy selection rule can be written as

$$\frac{SSE(\hat{\beta}_s)}{SSE(\hat{\beta}_{ls})} \leq 1 + \frac{Q}{K-Q-1} F_{Q,K-Q-1}(1-\alpha),$$

and concludes the collections of subsets obtained using R^2-adequacy are not larger than those subsets obtained using α-acceptability.

2.6.2.2 Multivariate Linear Regression Model

Basic Model. The multivariate linear regression model takes on the following form

$$Y_i = X_i\beta + \epsilon_i \ , \ 1 \leq i \leq K$$

where Y_i is the i^{th} independent observed response vector of dimension m, X_i is the i^{th} observed $1 \times (Q+1)$ row vector of predictors (with a 1 in the first column), β is a $(Q+1) \times m$ column vector of unknown parameters and ϵ_i is the i^{th} $1 \times m$ vector of residuals. We assume that $E(\epsilon_i) = \mathbf{0}$ and $\mathrm{cov}(\epsilon_i) = \Sigma$, $1 \leq i \leq K$. We note that

$$E[Y] = X\beta$$

and

$$\mathrm{var}[Y] = \Sigma,$$

if we can take X as fixed. As in the multiple regression case, we must be careful to distinguish between the case when X is fixed and the case when X is a random matrix. We can use conditioning arguments to show that we may estimate the parameters of the multivariate linear regression model using the same formulae for both random and fixed predictors. If we array all the K observations we have

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon$$

where $\mathbf{Y}$ is $K\times m$, $\mathbf{X}$ is $K\times(Q+1)$, β is $(Q+1)\times m$ and ϵ is $K\times m$. The least squares solution for β is

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y},$$

and an unbiased estimator of Σ for both methods is given as

$$\hat{\Sigma} = \frac{(\mathbf{Y} - \mathbf{X}\hat{\beta})'(\mathbf{Y} - \mathbf{X}\hat{\beta})}{K-Q-1}, \tag{2.6.4}$$

see Anderson (1984), theorem 8.2.1, pg. 289.

Selection of Variables

Relatively little has been written concerning the selection of predictor variables in the multivariate linear model context. Many papers have been written about variable selection in the closely related discriminant analysis problem; Seber (1984) provides a review.

Siotani, Hayakawa and Fujikoshi (1985) offer multivariate extensions to C_p and AIC. They also show how the forward selection technique can be extended to the multivariate case. Let RSS_p be the $m\times m$ matrix of residual sums of squares and cross-products,

$$RSS_p = (\mathbf{Y} - \mathbf{X}\hat{\beta})'(\mathbf{Y} - \mathbf{X}\hat{\beta}).$$

Siotani et al. generalize the C_p criterion to m responses as

$$C_p^m = \mathrm{tr}[\hat{\Sigma}^{-1} RSS_p] + 2mp - Km,$$

where $\hat{\Sigma}$ is as given in (2.6.4). They generalize AIC as

$$AIC^m = K \cdot \log \left| \frac{RSS_p}{K} \right| + 2mp.$$

To extend the forward selection technique, Siotani et al. suggest a generalization of the F-test using the Wilks Λ statistic. Let

$$\Lambda_p = \frac{|RSS_{p-1}|}{|RSS_p|},$$

in this case p-1 variables have been selected and the p^{th} variable is being considered as a candidate for entry. If the p^{th} variable affords a significant reduction in generalized residual variance then it will enter the model. Testing for nonsignificant reduction in generalized residual variance is equivalent to testing H_0: $\beta_p = \mathbf{0}$. Under H_0,

$$\frac{K-p-m}{m} \frac{1-\Lambda_p}{\Lambda_p} \sim F_{m,K-p-m},$$

so once again we perform F-tests on candidates for entry. If the candidate enters, we choose the next candidate by finding that predictor that, along with the p variables currently in the model, minimizes $|RSS_p|$.

McKay (1977) extends Aitkin's (1974) R^2-adequacy procedure to the multiresponse case. He explores the regressions between predictors and subsets of responses. This is an interesting approach, in that it allows one to think in terms of candidate responses as well as candidate predictors. McKay proposes several procedures, the most intuitive being a likelihood ratio-based simultaneous testing procedure. In this procedure selection is based on the squared canonical correlations between response and predictor subsets. This methodology, as well as those procedures due to Aitkin, are applications of the simultaneous test procedures offered by Gabriel (1969).

McKay reasons that the amount of information about the variation in a set of responses provided by a set of predictors is reflected in their squared canonical correlations. He shows that any subset of predictors can contain the same amount of information as the full set if and only if the slope *coefficients of the deleted variables are zero.* Let $\mathbf{Y}$, $\mathbf{X}$ *be jointly distributed* as

$$\begin{pmatrix} \mathbf{Y} \\ \mathbf{X} \end{pmatrix} \sim N_{m+Q} \left(\begin{pmatrix} \mu_Y \\ \mu_X \end{pmatrix}, \begin{pmatrix} \Sigma_{YY} & \Sigma_{YX} \\ \Sigma_{XY} & \Sigma_{XX} \end{pmatrix} \right),$$

further partition $\mathbf{X}' = (X'_f \ X'_g)$, $\mathbf{Y} = (Y'_v \ Y'_w)$, where $f \bigcup g = s$, $s = 1,\ldots,m$ and $v \bigcup w = u$, $u = 1,\ldots,Q$. This partitioning is arbitrary in the sense that we allow any pair of subsets to be represented. McKay's strategy is to apply one of Gabriel's simultaneous testing procedures to hypotheses of the form ω_{vs}: $\beta_{vs} = \mathbf{0}$ and of the form $\omega_{vg.f} = \mathbf{0}$, where $\omega_{vg.f}$ refers to a hypothesis in which the β matrix has already had those columns

corresponding to the partition g zeroed out. To get a simultaneous test with predetermined Type I family error rate, we test all such hypotheses versus the same critical value. The critical value arises from the "overall" hypothesis, that is that there does not exist a multivariate linear regression between $\mathbf{Y}$ and $\mathbf{X}$. A likelihood ratio-based simultaneous testing procedure is based on statistics of the following form,

$$\mathrm{W} = \frac{1 - \mathrm{U}}{\mathrm{U}}.$$

When the hypothesis is ω_{vs},

$$\mathrm{U}_{vs} = \frac{| S_{vv} - S_{vs} S_{ss}^{-1} S_{sv} |}{| S_{vv} |},$$

where S_{vv}, for instance, is the sample covariance matrix of the v responses. When the hypothesis is $\omega_{vg.f}$,

$$\mathrm{U}_{vg.f} = \frac{| S_{vv} - S_{vs} S_{ss}^{-1} S_{sv} |}{| S_{vv} - S_{vf} S_{ff}^{-1} S_{fv} |}.$$

We reject ω_{vs} when $\mathrm{U}_{vg.f} > \mathcal{W}^{\gamma}_{\mathrm{m,Q,K-Q-1}}$, where $\mathcal{W}^{\gamma}_{\mathrm{m,Q,K-Q-1}}$ is the γ-percentage point of the distribution of a random variable distributed as

$$\mathrm{V} = \frac{|\mathbf{G}|}{|\mathbf{G} + \mathbf{H}|},$$

where $\mathbf{G}$ is distributed as a Wishart distribution with $K - Q - 1$ degrees of

freedom, $\mathbf{H}$ is distributed as a Wishart distribution with Q degrees of freedom and $\mathbf{G}$ and $\mathbf{H}$ are independent. McKay defines γ-adequate subsets of predictors (X_f) for the subset of responses (X_v) as those $X_f : \mathbf{W}_{g.f} < w^{\gamma}_{\mathrm{m,Q,K-Q-1}}$.

CHAPTER 3
CONTROL VARIATE SELECTION CRITERIA

In this chapter we derive control variate selection criteria for two cases. First, we derive a criterion for the case when the covariance matrix of the controls is estimated. Next, we develop a new estimator that directly incorporates the use of a known covariance matrix for the controls. Finally, we present a selection criterion based on the new estimator.

3.1 A Selection Criterion When the Covariance Matrix of the Controls is Estimated

In this section we derive a criterion for use in the selection of control variates when the object is the immediate construction of confidence regions about the mean vector of responses. We demonstrate that this criterion acts as a multivariate extension of the univariate selection criterion, S_p.

Our objective is the immediate construction of controlled confidence regions. We use the word "immediate" to mean that we use only the data at hand, we do not make additional replications. It seems reasonable to pick that subset of controls which yields a confidence region of minimum expected volume. However, for mathematical efficacy, it is more convenient to consider minimizing the expected squared volume of the confidence region.

First, we will consider the case of a single response and show how the the appropriate criterion is a modified form of the S_p criterion. In the section following, we extend the criterion to multiple responses.

3.1.1 Univariate Response

In one dimension, the squared volume of the confidence region reduces to the squared width of the usual controlled confidence interval. Let $\mathbf{W}_\alpha^j$ correspond to the optimal width of a confidence interval constructed from $j \leq Q$ controls with a signifcance level α. We seek to find that subset of size j such that we minimize the expected squared length of the confidence interval: in symbols

$$\min_j E\left[(\mathbf{W}_\alpha^j)^2\right]. \tag{3.1.1.1}$$

Now

$$(\mathbf{W}_\alpha^j)^2 = 4t^2_{K-j-1}(1-\alpha/2)\hat{\sigma}_\epsilon^2\left\{s_{11}\right\}, \tag{3.1.1.2}$$

as developed in (2.2.39). We wish to compute

$$\min_j \; 4t^2_{K-j-1}(1-\alpha/2)E\left(\hat{\sigma}_\epsilon^2\left\{s_{11}\right\}\right) \tag{3.1.1.3}$$

but

$$E\left[\hat{\sigma}_\epsilon^2\left\{s_{11}\right\}\right] = \mathrm{var}(\overline{Y}(\hat{\beta})), \tag{3.1.1.4}$$

(by the arguments of (2.2.33) to (2.2.39)). Now, equation (2.2.30) gives

$$\mathrm{var}(\overline{Y}(\hat{\beta})) = \frac{\sigma_\epsilon^2}{K}\left[\frac{K-2}{K-j-2}\right]. \tag{3.1.1.5}$$

If we estimate σ_ϵ^2 by

$$\hat{\sigma}_\epsilon^2 = \frac{SSE_j}{K-j-1}, \tag{3.1.1.6}$$

then some algebra and recognition of those terms that are constant across all subsets yields a selection rule of

$$\min_j \; t^2_{K-j-1}(1-\alpha/2)\frac{SSE_j}{(K-j-1)(K-j-2)}, \tag{3.1.1.7}$$

but

$$\frac{SSE_j}{(K-j-1)(K-j-2)} = S_p \tag{3.1.1.8}$$

(actually S_p is S_j but we use p for the dimensionality of Y). The selection rule is

$$\min_j \; t^2_{K-j-1}(1-\alpha/2)S_p. \tag{3.1.1.9}$$

We note that if $K \gg j$ then the criterion reduces to S_p.

3.1.2 Multivariate Response

To extend this procedure to p responses, we seek the subset which yields the minimum expected squared volume of the confidence region. Rao (1967) gives, under the multivariate normal assumption, the 100(1−α)% confidence ellipsoid for μ_Y,

$$\Pr\left\{(\overline{\mathbf{Y}}(\hat{\beta})-\mu_{\mathrm{Y}})'\hat{\Sigma}_{Y|X}^{-1}(\overline{\mathbf{Y}}(\hat{\beta})-\mu_{\mathrm{Y}}) \leq (\mathbf{d}'\mathbf{d})C_0F_{p,K-Q-p}(1-\alpha)\right\} = 1-\alpha, \tag{3.1.2.1}$$

where

$$C_0 = [(K-Q-1)p/(K-Q-p)]$$

and

$$\mathbf{d}' = \mathbf{1}'_K/K - (\overline{\mathbf{X}}-\mu_{\mathrm{X}})'(\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}', \tag{3.1.2.2}$$

where G is defined in (2.4.25) and $\mathbf{1}_K$ is a K-dimensional column vector of ones. Also we have

$$\hat{\Sigma}_{\mathrm{Y}|\mathrm{X}} = \frac{K-1}{K-Q-1}(S_{\mathrm{YY}}-S_{\mathrm{YX}}S_{\mathrm{XX}}^{-1}S_{\mathrm{XY}}). \tag{3.1.2.3}$$

Let $(\mathrm{V}_\alpha^j)^2$ denote the squared volume of the confidence region constructed with $j \leq Q$ controls at significance level α. We seek to compute

$$\min_j E(\mathrm{V}_\alpha^j)^2. \tag{3.1.2.4}$$

Let $\tau_j = \left(\frac{p(K-j-1)}{K-j-p} \right)^p$; now, it can be shown that the volume of Rao's ellipsoid is given by

$$(V_\alpha^j)^2 = p^{-2}C^2(p)\,|\hat{\Sigma}_{Y|X}(\mathbf{d}'\mathbf{d})|\,\tau_j \left(F_{p,K-j-p}(1-\alpha) \right)^p, \qquad (3.1.2.5)$$

where $C^2(p)$ is given by (2.4.13). Rubinstein and Marcus (1985), pg. 675, calculate the expected value of $|\hat{\Sigma}_{Y|X}(\mathbf{d}'\mathbf{d})|$ as

$$|\Sigma_{Y|X}|\,C_3, \qquad (3.2.1.6)$$

where

$$C_3 = \prod_{i=1}^{p} \frac{K-j-i}{(K-j-1)K} \left(1 + \sum_{l=1}^{p} \binom{p}{l} \prod_{m=1}^{l} \frac{j+2(m-1)}{K-j-2m} \right), \qquad (3.1.2.7)$$

for $0 \leq K-j-2p$. So

$$E\left[(V_\alpha^j)^2\right] = p^{-2}C^2(p)\,|\Sigma_{Y|X}|\,C_3\tau_j \left(F_{p,K-j-p}(1-\alpha) \right)^p, \qquad (3.1.2.8)$$

and p is fixed for all subsets, hence we seek

$$\min_j \; |\Sigma_{Y|X}|\,C_3 \left(\frac{K-j-1}{K-j-p} \right)^p \left(F_{p,K-j-p}(1-\alpha) \right)^p, \qquad (3.1.2.9)$$

but

$$|\Sigma_{Y \cdot X}| = |\Sigma_{YY} - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}| = |I - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}\Sigma_{YY}^{-1}|\ |\Sigma_{YY}|$$

$$= |\Sigma_{YY}| \prod_{i=1}^{\nu}(1 - \rho_i^2), \qquad (3.1.2.10)$$

where $\nu_j =$ rank (Σ_{YX_j}). Now, $|\Sigma_{YY}|$ is not a function of j; and noting this and replacing the canonical correlations with estimates we get a criterion of

$$C_3\left(\frac{K-j-1}{K-j-p}\right)^p \left(F_{p,K-j-p}(1-\alpha)\right)^p \prod_{i=1}^{\nu_j}(1 - \hat{\rho}_i^2). \qquad (3.1.2.11)$$

3.2 A Selection Criterion When the Covariance Matrix of the Controls is Known

Situations arise in discrete event simulation where the covariance matrix of control variates is either known a priori or can be computed with relative ease. Several authors have suggested such controls for the class of closed queueing networks studied by Lavenberg et al. (1982). Wilson and Pritsker (198'a,b) and Venkatraman (1983) propose standardized control variables for these systems. In addition to having an asymptotically known mean, the controls offered by Wilson and Pritsker also have an asymptotically known covariance matrix. Venkatraman's controls have a mean vector known exactly with a covariance matrix that is also known asymptotically. In Chapter 4 of this research, we offer a new class of controls for which both the mean vector and covariance matrix are known asymptotically. To emphasize

the potential diversity of situations where the covariance matrix is known, we cite the so-called "path" controls offered by Venkatraman and Wilson (1985). These controls arise in the simulation of stochastic activity networks. We believe that there is a large class of simulated systems for which such controls can be developed.

In this section we develop a controlled estimator $\overline{Y}(\hat{\gamma})$ that directly incorporates the known covariance matrix of the controls. We derive its various properties and develop a selection criterion based on this estimator.

3.2.1 The Estimator $\overline{Y}(\hat{\gamma})$

In Section 2.4 we introduced the estimator

$$\overline{Y}(\hat{\beta}) = \overline{\mathbf{Y}} - \hat{\beta}(\overline{\mathbf{X}}-\mu_{\mathbf{X}}) \;, \tag{3.2.1.1}$$

where $\hat{\beta} = S_{YX}S_{XX}^{-1}$. Here, as in Section 2.4, $\mathbf{Y}$ is a $p \times 1$ vector of responses, $\hat{\beta}$ is the estimated $p \times Q$ matrix of control coefficients and $\mathbf{X}$ is a $Q \times 1$ vector of controls with mean vector $\mu_{\mathbf{X}}$. In this section we will introduce the estimator

$$\overline{\mathbf{Y}}(\hat{\gamma}) = \overline{\mathbf{Y}} - \hat{\gamma}(\overline{\mathbf{X}}-\mu_{\mathbf{X}}) \tag{3.2.1.2}$$

where

$$\hat{\gamma} = S_{YX}\Sigma_{XX}^{-1} \tag{3.2.1.3}$$

This estimator incorporates a known covariance matrix, Σ_{XX}, for the controls.

Provided that the responses and controls are jointly normal, $\overline{\mathbf{Y}}(\hat{\gamma})$ is an unbiased estimator for μ_Y. To show this we write $\overline{\mathbf{Y}}(\hat{\gamma})$ as a linear combination of the K observed responses. Let

$$\mathbf{G} = \begin{bmatrix} (X_1 - \overline{X})' \\ \cdot \\ \cdot \\ \cdot \\ (X_K - \overline{X})' \end{bmatrix} \tag{3.2.1.4}$$

and define

$$\tilde{\mathbf{H}} = K^{-1}\mathbf{1}_K - (K-1)^{-1}\mathbf{G}\Sigma_{XX}^{-1}(\overline{\mathbf{X}}-\mu_{\mathbf{X}}) \ . \tag{3.2.1.5}$$

We observe

$$\mathbf{1}_K{}'\mathbf{G} = \mathbf{0}, \ \ \mathbf{1}_K{}'\mathbf{H} = 1, \ \ \begin{bmatrix} Y_1, \cdots, Y_K \end{bmatrix}\mathbf{G} = (K-1)S_{YX} \ , \tag{3.2.1.6}$$

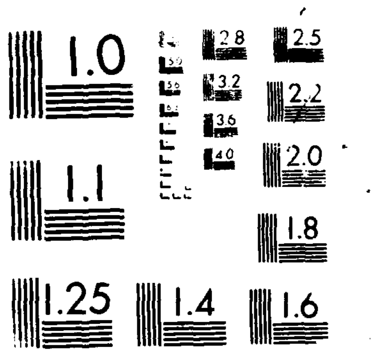
and so we write

$$\overline{\mathbf{Y}}(\hat{\gamma}) = \begin{bmatrix} Y_1, \cdots, Y_K \end{bmatrix}\tilde{\mathbf{H}} \ . \tag{3.2.1.7}$$

Now, define $\mathbf{Z} =$ vec $\mathbf{X} =$ vec $(X_1, \cdots, X_K)$ so that $\mathbf{Z}$ is the KQ- dimensional column vector formed by stacking the Q-dimensional [illegible] $\{X_j\}$ one upon another. Now, the condition $\{X_j = x_j\}$ [illegible]

AD-A186 637 CONTROL VARIATE SELECTION FOR MULTIRESPONSE SIMULATION 2/3
(U) AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH
K W BAUER MAY 87 AFIT/CI/NR-87-132D

UNCLASSIFIED F/G 12/3 NL



RESOLUTION TEST CHA

NATIONAL BUREAU OF STANDARDS

compactly expressed as $\mathbf{Z} = z$. Also, define $\mu_Z = \text{vec}\,(\mathbf{1}_K \otimes \mu_X)$. We have

$$E\left[\bar{\mathbf{Y}}(\hat{\gamma})\right] = E_X\left[E_Y\left[\text{ vec } \mathbf{Y}\tilde{\mathbf{H}} \,|\, \mathbf{Z} = z\right]\right] \tag{3.2.1.8}$$

$$= E_X\left[(\tilde{\mathbf{H}}' \otimes I_p) E_Y\left[\text{ vec } \mathbf{Y} \,|\, \mathbf{Z} = z\right]\right]$$

Now, assuming joint normality of the responses and controls, we have

$$= E_X\left[(\tilde{\mathbf{H}}' \otimes I_p)\left[(\mathbf{1}_K \otimes \mu_Y) + (I_K \otimes \Sigma_{YX}\Sigma_{XX}^{-1})(\mathbf{Z} - \mu_Z)\right]\right]$$

$$= E_X\left[(1 \otimes \mu_Y) + (\tilde{\mathbf{H}}' \otimes \Sigma_{YX}\Sigma_{XX}^{-1})(\mathbf{Z} - \mu_Z)\right]$$

$$= E_X\left[\mu_Y + (\tilde{\mathbf{H}}' \otimes \Sigma_{YX}\Sigma_{XX}^{-1})\mathbf{Z} - (\tilde{\mathbf{H}}' \otimes \Sigma_{YX}\Sigma_{XX}^{-1})(\mathbf{1}_K \otimes \mu_X)\right]$$

$$= E_X\left[\mu_Y + (\Sigma_{YX}\Sigma_{XX}^{-1}\mathbf{X}\tilde{\mathbf{H}}) - (1 \otimes \Sigma_{YX}\Sigma_{XX}^{-1}\mu_X)\right]$$

$$= \mu_Y + E_X(\Sigma_{YX}\Sigma_{XX}^{-1}\mathbf{X}\tilde{\mathbf{H}}) - (\Sigma_{YX}\Sigma_{XX}^{-1}\mu_X)$$

$$= \mu_Y + \Sigma_{YX}\Sigma_{XX}^{-1}E_X\left[\mathbf{X}\tilde{\mathbf{H}}\right] - (\Sigma_{YX}\Sigma_{XX}^{-1}\mu_X)$$

$$= \mu_Y + \Sigma_{YX}\Sigma_{XX}^{-1}E_X\left[\bar{\mathbf{X}} - (K-1)^{-1}\mathbf{XG}\Sigma_{XX}^{-1}(\bar{\mathbf{X}}-\mu_{\mathbf{X}})\right] - (\Sigma_{YX}\Sigma_{XX}^{-1}\mu_X)$$

$$= \mu_Y + E_X\left[\mathbf{S}_{XX}\Sigma_{XX}^{-1}(\bar{\mathbf{X}}-\mu_{\mathbf{X}})\right],$$

and under the assumption of multivariate normality, we have

$$= E_X\left[\mathbf{S}_{XX}\Sigma_{XX}^{-1}\right]E_X\left[(\bar{\mathbf{X}}-\mu_{\mathbf{X}})\right] =$$

$$E_X\left[\mathbf{S}_{XX}\Sigma_{XX}^{-1}(\bar{\mathbf{X}}-\mu_{\mathbf{X}})\right] = \mathbf{0} \tag{3.2.1.9}$$

since $\mathbf{S}_{XX}$ is independent of $\bar{\mathbf{X}}$ in this case. Hence the estimator, $\bar{\mathbf{Y}}(\hat{\gamma})$, is unbiased. The covariance of $\bar{\mathbf{Y}}(\hat{\gamma})$ is given by

$$\operatorname{cov}\left[\bar{\mathbf{Y}}(\hat{\gamma})\right] = \left[\frac{K+Q-1}{K(K-1)}\right]\Sigma_{Y|X} + \left[\frac{Q+1}{K(K-1)}\right]\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY} \tag{3.2.1.10}$$

The derivation of (3.2.1.10) is Appendix 1. Algebraic manipulation of the above reveals

$$\operatorname{cov}\left[\bar{\mathbf{Y}}(\hat{\gamma})\right] = \frac{K+Q-1}{K(K-1)}\Sigma_{YY}\left[I_p - \frac{K-2}{K+Q-1}\Sigma_{YY}^{-1}\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}\right], \tag{3.2.1.11}$$

hence the variance ratio is given by

$$\eta_2^* = \frac{\det(\mathrm{cov}(\overline{\mathbf{Y}}(\hat{\gamma})))}{\det(\mathrm{cov}(\overline{\mathbf{Y}}))} \tag{3.2.1.12}$$

$$= \left(\frac{K+Q-1}{K-1}\right)^p \prod_{i=1}^{\nu} \left(1 - \frac{K-2}{K+Q-1}\rho_i^2\right), \tag{3.2.1.13}$$

where $\nu =$ rank Σ_{YX}. Let ρ_* be the smallest canonical correlation, then

$$\eta_2^* < \left(\frac{K+Q-1}{K-1}\right)^p \left(1 - \frac{K-2}{K+Q-1}\rho_*^2\right)^{\nu}. \tag{3.2.1.14}$$

Now if $\nu = p$, a resonable condition corresponding to well-picked responses and at least p well-picked controls, then the right hand side of (3.2.1.14) can be shown to be less than 1 (see Appendix 2) if

$$\rho_* > \left(\frac{Q}{K-2}\right)^{1/2} \tag{3.2.1.15}$$

We have shown that it is possible to achieve a reduction in the generalized variance using $\overline{\mathbf{Y}}(\hat{\gamma})$ as an estimator. However, for $\overline{\mathbf{Y}}(\hat{\gamma})$ to be of practical significance we require it to reduce the generalized variance relative to its competitor $\overline{Y}(\hat{\beta})$. To contrast the estimators $\overline{\mathbf{Y}}(\hat{\gamma})$ and $\overline{Y}(\hat{\beta})$, we form the ratio

$$\tilde{\eta} = \frac{\det(\mathrm{cov}(\overline{\mathbf{Y}}(\hat{\gamma})))}{\det(\mathrm{cov}(\overline{Y}(\hat{\beta})))} = \frac{\left(\frac{K+Q-1}{K-1}\right)^{p} \prod_{i=1}^{\nu} \left(1 - \frac{K-2}{K+Q-1}\rho_i^2\right)}{\left(\frac{K-2}{K-Q-2}\right)^{p} \prod_{i=1}^{\nu} \left(1 - \rho_i^2\right)}$$

(3.2.1.16)

Using mathematical induction and assuming that $\nu = p$ (see Appendix 3), we can show

$$\rho_*^2 > \frac{\left(\frac{(K+Q-1)(K-Q-2)}{(K-1)(K-2)}\right) - 1}{\left(\frac{(K+Q-1)(K-Q-2)}{(K-1)(K-2)}\right)\left(\frac{K-2}{K+Q-1}\right) - 1} \Rightarrow \tilde{\eta} < 1$$

(3.2.1.17)

We have shown, subject to conditions on the canonical correlations, that reduction in the generalized variance of the estimator is possible. The next step is to create a 100(1$-\alpha$)% confidence region based on $\overline{\mathbf{Y}}(\hat{\gamma})$.

We construct a 100(1$-\alpha$)% confidence region based on the following considerations. We assume

$$\overline{\mathbf{Y}}(\hat{\gamma}) \sim N_p(\mu_Y, \tilde{\Sigma}) , \qquad (3.2.1.18)$$

where $\tilde{\Sigma}$ is given by (3.2.1.10). Let $\hat{\tilde{\Sigma}}$ be an estimator of $\tilde{\Sigma}$, further; assume

$$(n-Q)\hat{\tilde{\Sigma}} \sim \mathcal{W}_p(n-Q \ , \tilde{\Sigma}) , \qquad (3.2.1.19)$$

where $\mathcal{W}_p(n-Q\ ,\ \tilde{\Sigma})$ is a random matrix distributed as Wishart with $n-Q$ degrees of freedom and expected matrix $\tilde{\Sigma}$; here $n = K-1$. Now if $\bar{Y}(\hat{\gamma})$ is independent of $\hat{\hat{\Sigma}}$ and we define

$$T^2 = (\bar{Y}(\hat{\gamma})-\mu_Y)'\hat{\hat{\Sigma}}^{-1}(\bar{Y}(\hat{\gamma})-\mu_Y)\ , \tag{3.2.1.20}$$

we can show (see Muirhead (1982), Theorem 3.2.13, pg. 98)

$$\frac{T^2}{K-Q-1}\,\frac{K-Q-p}{p} \sim F_{p,K-p-Q}\ , \tag{3.2.1.21}$$

where $F_{p,K-p-Q}$ is a random variable distributed as a central F with p and $K-p-Q$ degrees of freedom respectively. We can form a confidence region based on

$$\Pr\left\{(\bar{Y}(\hat{\gamma})-\mu_Y)'\hat{\hat{\Sigma}}^{-1}(\bar{Y}(\hat{\gamma})-\mu_Y) \le C_0 F_{p,K-Q-p}(1-\alpha)\right\} = 1-\alpha, \tag{3.2.1.22}$$

where

$$C_0 = [(K-Q-1)p\,/(K-Q-p)]$$

3.2.2 A Selection Criterion

As in the previous section, our objective is the immediate construction of controlled confidence regions. Again, we will seek that subset of controls that yields a confidence region of minimum expected squared volume. Expression (3.2.1.22) is used to construct the $100(1-\alpha)\%$ confidence ellipsoid for μ_Y. Let $(\mathrm{V}_\alpha^j)^2$ correspond to the optimal squared volume of the confidence region constructed with $j \leq Q$ controls at significance level α. We seek

$$\min_j E(\mathrm{V}_\alpha^j)^2. \tag{3.2.2.1}$$

Let $\tau_j = \left(\frac{p(K-j-1)}{K-j-p}\right)^p$; now, it can be shown that the squared volume of this ellipsoid is given by

$$(\mathrm{V}_\alpha^j)^2 = p^{-2}C^2(p)\,|\hat{\hat{\Sigma}}_j|\,\tau_j \left(F_{p,K-j-p}(1-\alpha)\right)^p, \tag{3.2.2.2}$$

where $C(p)$ is given by (2.4.13). Assume $(K-1-j)\hat{\hat{\Sigma}}_j \sim \mathcal{W}_\mathrm{p}(\mathrm{K}-1-\mathrm{j}\,,\,\tilde{\Sigma})$. Here $\hat{\hat{\Sigma}}_j$ depends on the subset of j controls. Using Theorem 3.2.15 of Muirhead (1982), we calculate the expected value of $|\hat{\hat{\Sigma}}_j|$ as

$$|\tilde{\Sigma}_j|\,C_4\,, \tag{3.2.2.3}$$

where

$$C_4 = \prod_{i=1}^{p} \frac{K-j-i}{K-j-1}$$

for $K-j-1 > 0$. So

$$E\left[(\mathrm{V}_{\alpha}^{j})^2\right] = p^{-2}C^2(p)\,|\tilde{\Sigma}_j|\,C_4\tau_j\left(F_{p,K-j-p}(1-\alpha)\right)^p, \qquad (3.2.2.4)$$

and since p is fixed for all j, we seek

$$\min_j \;|\tilde{\Sigma}_j|\,C_4\left(\frac{K-j-1}{K-j-p}\right)^p\left(F_{p,K-j-p}(1-\alpha)\right)^p. \qquad (3.2.2.5)$$

We do not know $\tilde{\Sigma}_j$ so we estimate it as

$$\left[\frac{K+j-1}{K(K-1)}\right]\hat{\Sigma}_{Y\,|\,X} + \left[\frac{j+1}{K(K-1)}\right]\left(\hat{\Sigma}_{YY} - \hat{\Sigma}_{Y\,|\,X}\right), \qquad (3.2.2.6)$$

where

$$\hat{\Sigma}_{Y\,|\,X} = \frac{K-1}{K-j-1}\left(S_{YY} - S_{YX}S_{XX}^{-1}S_{XY}\right), \qquad (3.2.2.7)$$

and

$$\hat{\Sigma}_{YY} = S_{YY}\;, \qquad (3.2.2.8)$$

where S_{YY} is defined in equation (2.4.17).

CHAPTER 4
IMPLEMENTATION OF THE SELECTION CRITERIA IN QUEUEING NETWORK SIMULATION

In this chapter we discuss the evaluation portion of the research. First, we describe a class of closed queueing networks that were used as the experimental vehicles of this research. Then, we describe an experimental procedure in which a series of simulation metaexperiments were carried out to evaluate the performance of the multivariate selection criteria. Here we discuss the system responses investigated and the candidate controls. Next, we discuss the the performance measures employed. Finally, we discuss the methodology used to obtain the optimal subset of controls in each basic experiment.

4.1 Description of the Simulated Queueing Networks

We chose four different queueing systems as experimental vehicles. These queueing systems were suggested by Lavenberg et al. (1982). These systems are members of a broad class of closed queueing networks. There are several major advantages to choosing such systems. The first major advantage lies in the fact that these systems have been studied extensively and workable controls have been developed by several authors. Lavenberg et al. (1982) have developed a set of control variables that, in the univariate

response case, have produced significant variance reductions. Wilson and Pritsker (1984a, b) show how to modify these controls to insure asymptotic stability. Venkatraman (1983) offers a standardization scheme for one subset of the controls of Lavenberg et al. Another advantage in using these networks is that they are representative of a class of queueing systems which are frequently analyzed in computer performance modeling. Finally, the two simpler networks we employ have a steady-state behavior which can obtained analytically. This information was of great value for validation.

As outlined in Lavenberg et al., the queueing systems considered take the following form. Consider a finite set (say of size S) of interconnected service centers. These centers service D different types of customers. There are a total of N customers of all types. Assume

1. Markovian routing so that the next station visited only depends on the current location.

2. The service times for the the j^{th} type of customer at the i^{th} service station are drawn independently from a given distribution $F_{ij}(*)$ with finite mean and variance.

3. Service time sequences and sequences of centers visited are mutually independent.

There are two basic types of networks to be considered in this general setting. Figure 1 portrays the form of the first type of simulated network. Service center 1 has N servers, where N is the total number of customers of all types. We can think of this service center as a room filled with N interactive computer terminals. The service centers labeled $2, \cdots, S$ are

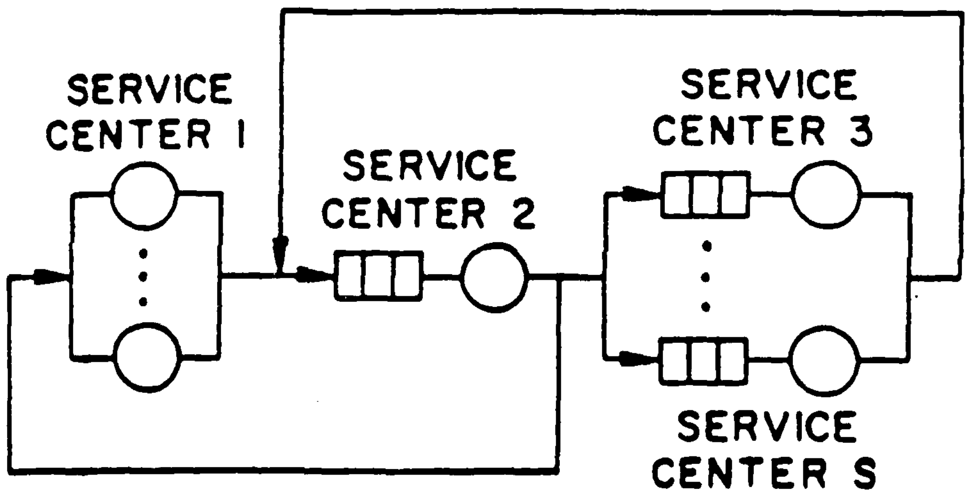


Figure 1. Type I Network

single server queues with the customers being served in order of arrival. We can think of service center 2 as a central processing unit (CPU) with service centers 3, $\cdots$, S as peripheral devices to be accessed by the CPU. The S by S transition matrix that characterizes the flow of customers in the network has the form

$$\mathrm{P}(d) = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ p_1(d) & 0 & p_2(d) & p_3(d) & \cdots & p_S(d) \\ 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 \end{bmatrix},$$

where $p_k(d)$ $k = 1, \cdots, S$ is the one step transition probability from service center 2 to the remaining centers (for a customer of type d). In this type of network we have made the implicit assumption that every customer that requests service from the CPU is immediately granted his requisite memory allocation. In real world interactive computing enviroments, customers often must queue for memory at the CPU. This blocking effect due to memory limitations of the CPU is modelled by the next type of network.

We refer to this second class of systems as networks with subnetwork capacity constraints. The CPU and associated peripherals are the subnetwork. A network of this type is portrayed in Figure 2. The dotted line encloses the subnetwork. Service center 2 is now merely a queue for the subnetwork with capacity $N' < N$ customers. There is no service time associated with service center 2.

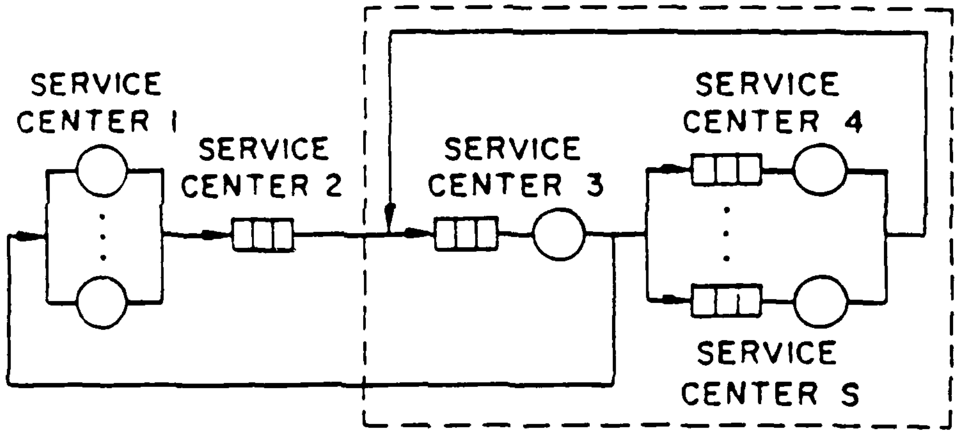


Figure 2. Type II Network

In this case the S by S transition matrix takes the form

$$P(d) = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ p_1(d) & 0 & 0 & p_2(d) & \cdots & p_S(d) \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \end{bmatrix}.$$

In our experiments we chose 4 networks from Lavenberg et al. (1982). Two of the networks were of the first general type, i.e., no subnetwork capacity constraints. The other two networks had subnetwork capacity constraints. The networks are parameterized in the tables below.

Table 4.1 Parameters of Queueing Systems Used in the Experimental Evaluation

Network No.	N No. customers	Subnetwork Capacity N'	S Number of Service Centers
1	25	25	4
2	15	15	4
3	25	5	7
4	25	10	7

Table 4.2 Mean Service Times for the Queueing Systems Used in the Experimental Evaluation

Network No.	Service Center Number						
	1	2	3	4	5	6	7
1	100	1	.694	6.25	—	—	—
2	100	1	2.78	25.0	—	—	—
3	100	—	1	2.78	2.78	25	25
4	100	—	1	2.78	2.78	25	25

Table 4.3 Branching Probabilities for the Queueing Systems Used in the Experimental Evaluation

Network No.	Probability of Branching from Central Server To Station j						
	1	2	3	4	5	6	7
1	.2	0	.72	.08	—	—	—
2	.2	0	.72	.08	—	—	—
3	.2	0	0	.36	.36	.04	.04
4	.2	0	0	.36	.36	.04	.04

We chose the first two networks because Lavenberg et al. presented detailed results of their experiments using these two networks. Also, these two simpler networks display a steady state behavior that can be obtained analytically. Results for the networks 3 and 4 were presented in Lavenberg, Moeller, and Welch (1979).

4.2 Layout of the Simulation Experiments

In this section we discuss elements of the experimental layout. First, we briefly describe the relationship of our basic experiment to an overall metaexperiment. In the next section, we discuss the selected system responses. Following that, we list the selected control variables. Finally, we discuss the selected performance measures.

4.2.1. Composition of the Metaexperiments

A basic experiment consisted of a set number of independent replications of the simulation model. We chose two replication levels, 20 and 40. Within a basic experiment a selection procedure was employed to obtain the "best" subset of controls. We discuss the selection procedure in a later section. An overall metaexperiment was conducted for each network. This meta-experiment consisted of 50 independent replications of the basic experiment. Therefore, when the replication level of a basic experiment was 40, we ran $40 \times 50 = 2000$ independent replications of the basic experiment.

4.2.2. Selected System Responses

We elected to look at a two dimensional response across all the models. We were interested in the response vector $\left(R(t), U_{CPU}(t)\right)$ where $R(t)$ is the mean response time for a system request accumulated up to simulated time t and $U_{CPU}(t)$ is the utilization of the CPU accumulated up to simulated time t. These responses seem to make good sense from both a system and customer viewpoint. The customer is most interested in the response time of the system and the system adminstrator is probably most concerned with the utilization on the CPU, since it is probably by far the most expensive component of his system.

4.2.3. Selected Control Variables

We considered, as candidates, modifed versions of the control variables proposed by Lavenberg et al. (1982), as well as as new control of our own design. These control variables can be classified into three basic types: 1) service time variables, 2) flow variables, and 3) work variables. All of these variables are collected at each service center for each customer type. In the form suggested by Lavenberg et al., service time variables are the sample mean service times. Flow variables are the sample proportion of departures from particular centers relative to the total number of departures from all centers. Work variables are the product of the service time variables and the flow variables.

Lavenberg et al. calculated the asymptotic means for these controls. In their analysis, they assumed that the run lengths of their models were sufficient to warrant the use of the asymptotic means. We chose to use

"standardized" forms of the service and work variables of Lavenberg et al. We developed a standardized flow-type variable based on the multinomial routing of each network. We used "standardized" controls in an effort to avoid numerical difficulties that result from unit of measurement differences.

We considered, as candidates, service time variables of the type proposed by Venkatraman (1983). For service center j and customer type d, define

$$X_{j,d}(t) = (t)^{-1/2} \sum_{l=1}^{g(j,d;t)} (U_{j,d,l}(d) - \mu_{j,d})/\sigma_{j,d} ,$$

where: $g(j, d; t)$ is the number of service times started at station j for customer class d during the simulated interval $(0, t]$ (here 0 marks the start of the statistics collection period); and $U_{j,d,\,l}$ is the l^{th} service time sampled at station j for customer class d, where $E(U_{j,d,\,l}) = \mu_{j,d}$ and $\mathrm{var}(U_{j,d,\,l}) = \sigma^2_{j,d}$.

Venkatraman (1983) shows that

$$E[\mathbf{X}(t)] \equiv \mathbf{0} \ \forall \ t \geq 0$$

where

$$\mathbf{X}(t) = [X_1(t), \ \cdots \ , X_q(t)]$$

and since

$$\mathbf{X}(t) \xrightarrow[t \to \infty]{D} N(\mathbf{0}, \alpha)$$

where $\xrightarrow[t \to \infty]{D}$ signifies convergence in distribution and

$$\alpha = \text{diag}(\alpha_1, \cdots, \alpha_q)$$

and

$$\alpha_k = (\text{Steady-state utilization of station } k)/\mu_k$$

(note that we have dropped the customer class distinction). Venkatraman provides guidance for the computation of α_k.

Among the candidate control variables are versions of the work variables due to Lavenberg et al. that have been suitably "standardized" as proposed by Wilson and Pritsker (1984a, b) (see equation 2.2.48). This class of controls are attractive because the vector of q standardized work variables converges to a q-variate normal distribution with a zero mean vector and unit dispersion matrix:

$$X^*(t) \xrightarrow[t \to \infty]{D} N_q(\mathbf{0}, \mathbf{I}) .$$

4.2.4 Routing Control Variables

The moments of the flow variables proposed by Lavenberg et al. are in general unknown. Hence, we were not able to standardize these variables. We discarded these controls as candidates in favor of a standardized multinomial control. We call these new controls, "routing variables".

All routing in the networks we considered is done from the CPU. Now define an indicator variable on the event of the l^{th} departure from the CPU to station j

$$U_l(j) = \begin{cases} 1 \text{ if the } l^{th} \text{ departing customer goes to station j} \\ 0 \text{ otherwise} \end{cases}$$

Now, from the discussion of the simulated networks we have $p_j(*)$ as the probabilty of transition from the CPU to station j. If $N(t)$ is the total number of transitions from the CPU up to time t then

$$\sum_{l=1}^{N(t)} U_l(j) \mid N(t) = n \sim B(n; p_j(*)) ,$$

a binomial random variable on n trials with success probability $p_j(*)$. We consider standardized controls of the form

$$X_j^R(t) = \begin{cases} \sum_{l=1}^{N(t)} \dfrac{U_l(j) - p_j(*)}{\sqrt{N(t)(1-p_j(*))p_j(*)}} & \text{for} \quad j = 1, \cdots, S \text{ if } N(t) > 0 \\ 0 & \text{if } N(t) = 0 \end{cases}$$

Since $E\left[X_j^R(t) \mid N(t)\right] = 0$ for all n and for all t, we see that $X_j^R(t)$ has mean

zero not merely asymtotically but for all times t:

$$E\left[X_j^R(t)\right] = 0 \quad \text{for all } t \geq 0 \text{ ;}$$

moreover

$$\text{var}\left[X_j^R(t) \,|\, N(t) = n\right] = 1 \quad \text{for } n > 0 \Rightarrow \lim_{t \to \infty} \text{var}\left[X_j^R\right] = 1 \text{ .}$$

The covariance matrix for the routing variables is given by the matrix $\Sigma_{RR} = \lim_{t \to \infty} \left[\text{cov}\left(X_j^R(t), X_k^R(t)\right) | \, N(t)\right]$, where j and k reflect possible routings from the CPU. The covariance matrix is calculated as

$$\text{cov}\left[X_j^R(t), X_k^R(t) \,|\, N(t)\right] =$$

$$\begin{cases} 1 & \text{when } j{=}k \text{ , } N(t){>}0 \\[2ex] \dfrac{-\sqrt{p_j(^*)p_k(^*)}}{\sqrt{(1 - p_j(^*))(1 - p_k(^*))}} & \text{when } j{\neq}k \text{ , } N(t){>}0 \\[2ex] 0 & \text{when } N(t){=}0 \end{cases}$$

(4.2.4.1)

The derivation of (4.2.4.1) is given in Appendix 4.

Standardized routing variables share the same desirable asymptotic properties as the standardized controls proposed by Wilson and Pritsker (1984a, b). Let $X^R(t) \equiv \left[X_1^R(t), \ldots, X_s^R(t)\right]$. In Appendix 5 we show that

$$X^R(t) \xrightarrow[n \to \infty]{D} N_s(\mathbf{0}, \Sigma_{RR}) , \tag{4.2.4.2}$$

in a broad class of queueing networks that possess the regenerative property. We require $s = S - 1$. This requirement is explained in Appendix 5.

4.2.5 Selected Performance Measures

To assess the efficiency of using the "best" subset of controls, we concentrated on two performance measures: (a) coverage and (b) volume reduction. Coverage was computed, across a metaexperiment, as the actual proportion of generated confidence ellipsoids that contain the true mean vector. Volume reduction was computed, across a metaexperiment, as the average percentage reduction in the volume of the confidence ellipsoid generated with the "best" subset of controls relative to the volume of the ellipsoid generated by direct simulation.

For experiment l, let

$$\hat{V}_l^* = \begin{cases} \text{Volume of the confidence ellipsoid} \\ \text{generated with the best subset of} \\ \text{controls,} \end{cases}$$

$$\hat{V}_l^d = \begin{cases} \text{Volume of the confidence ellipsoid} \\ \text{generated by direct simulation,} \end{cases}$$

$$\hat{P}_l^{\prime} = \begin{cases} 1 & \text{if the controlled confidence ellipsoid} \\ & \text{contains the true mean vector} \\ 0 & \text{otherwise,} \end{cases}$$

$$\hat{P}_l^{d} = \begin{cases} 1 & \text{if the uncontrolled confidence ellipsoid} \\ & \text{contains the true mean vector} \\ 0 & \text{otherwise,} \end{cases}$$

Across the metaexperiment we compute

$$\hat{V}^{\prime} = \frac{1}{50} \sum_{l=1}^{50} \hat{V}_l^{\prime}$$

$$\hat{V}^{d} = \frac{1}{50} \sum_{l=1}^{50} \hat{V}_l^{d}$$

$$\hat{P}^{\prime} = \frac{1}{50} \sum_{l=1}^{50} \hat{P}_l^{\prime}$$

$$\hat{P}^{d} = \frac{1}{50} \sum_{l=1}^{50} \hat{P}_l^{d}$$

so that the final performance measures are

$$\text{Volume Reduction (\%)} = \left[1 - \frac{\hat{V}^{\prime}}{\hat{V}^{d}} \right] \times 100$$

$$\begin{array}{l}\text{Coverage Probabilty}(\%)\\ \text{(Direct Simulation)}\end{array} = \hat{\mathrm{P}}^{d} \times 100$$

$$\begin{array}{l}\text{Coverage Probabilty}(\%)\\ \text{(Best Subset of Controls)}\end{array} = \hat{\mathrm{P}}^{\prime} \times 100$$

When $\overline{Y}(\hat{\beta})$ is used, the expression used for volume reduction can be obtained from equations 34 and 35 of Rubinstein and Marcus (1985). Specifically

$$\frac{\hat{\mathrm{V}}^{*}}{\hat{\mathrm{V}}^{d}} = \frac{|\hat{\Sigma}_{Y \cdot X}|^{1/2}}{|S_{YY}|^{1/2}} (\mathrm{d}'\mathrm{d})^{p/2} \left[\frac{(K-Q-1)K(K-p)}{(K-1)(K-Q-p)} \right]^{p/2} \left[\frac{F_{p,K-Q-p}(1-\alpha)}{F_{p,K-p}(1-\alpha)} \right]^{p/2}$$

and when $\overline{Y}(\hat{\gamma})$ is used, we have

$$\frac{\hat{\mathrm{V}}^{*}}{\hat{\mathrm{V}}^{d}} = \frac{|\hat{\hat{\Sigma}}|^{1/2}}{|S_{YY}|^{1/2}} \left[\frac{(K-Q-1)K(K-p)}{(K-1)(K-Q-p)} \right]^{p/2} \left[\frac{F_{p,K-Q-p}(1-\alpha)}{F_{p,K-p}(1-\alpha)} \right]^{p/2},$$

where all notation is as appears in Section 2.4. After some pilot experimentation we decided to employ the standardized work variables of Wilson and Pritsker (1984a, b), as well as the standardized routing variables proposed above.

4.3 Optimal Subset Selection Methodology

Within each basic experiment, a set number of replications of the simulation model were performed. Response and control variable data were collected for each replication. After the data had been collected for the last replication within the basic experiment, a control variable selection methodology was applied against the data. This procedure computes the multivariate selection criterion for all possible subsets of controls and finds

the minimum. The subset corresponding to the minimum value of the selection criterion is deemed "best" or "optimal". Given the optimal set of controls, the coverage and confidence region volume reduction measures were computed and tallied.

The selection procedure is initiated by the construction of a grand covariance matrix that includes the responses and the controls. Next, a multivariate generalization of a binary search of the regression "tree", as proposed by Furnival and Wilson (1974), is employed to examine all possible subsets of controls. Finally, the procedure computes the performance statistics.

As demonstrated in Sections 2.1 and 2.2, the control variate technique can be viewed as a linear regression problem. In Section 2.1, we consider the case where there is one response and one control. In this case, if we assume joint normality for Y and X, then conditional on X=x, we have the classical regression problem

$$\mathbf{Y} = \mathbf{D}\alpha + \epsilon \,, \tag{4.3.1.1}$$

where $\mathbf{Y} = (Y_1, \cdots, Y_K)$, $\alpha = \begin{pmatrix} \mu_Y \\ \beta \end{pmatrix}$, $\mathbf{D}$ is as in equation (2.1.17), K is the number of replications, $\epsilon' = (\epsilon_1, \cdots, \epsilon_K)$ a vector of residuals such that $\epsilon_i \sim \text{IID } N(0, \sigma_\epsilon^2)$, and β is as in equation (2.1.11).

Under the multinormal assumption, the least squares and maximium likelihood solution for α is given by

$$\hat{\alpha} = (\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'Y \,.$$

We note that the only remaining unknown in the model is the variance of the residuals. An unbiased estimator is given by

$$\hat{\sigma}_\epsilon^2 = \frac{(Y - \mathbf{D}\hat{\alpha})'(Y - \mathbf{D}\hat{\alpha})}{K-2} \ .$$

It is apparent that the computational overhead necessary to completely estimate the model of equation (4.3.1.1) lies mostly in the inversion of $\mathbf{D}'\mathbf{D}$.

When we extend the model of equation (4.3.1.1) to the case of a univariate response with multiple controls, as well as to the case of multiple responses with multiple controls, we see that little changes from a computational viewpoint. In particular, for the case of a univariate response with multiple controls, the model becomes

$$\mathbf{Y} = \mathbf{D}\alpha + \epsilon \ , \tag{4.3.1.2}$$

where $\mathbf{Y} = (Y_1, \cdots, Y_K)$, $\alpha' = (\mu_Y, \beta_1, \cdots, \beta_q)$, $\mathbf{D}$ is lescribed by equation (2.2.35), K is the number of replications, $\epsilon' = (\epsilon_1, \cdots, \epsilon_K)$ is a vector of residuals such that $\epsilon_i \sim$ IID $N(0, \sigma_\epsilon^2)$, β_i is as in equation (2.2.33).

Under the multinormal assumption, the least-squares and maximum likelihood estimates for α and σ_ϵ^2 are respectively given by

$$\hat{\alpha} = (\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'Y \ .$$

$$\hat{\sigma}_\epsilon^2 = \frac{(Y - \mathbf{D}\hat{\alpha})'(Y - \mathbf{D}\hat{\alpha})}{K-Q-1}$$

Seber (1977). Finally, in the case of multiple responses and multiple controls, the model is

$$\mathbf{Y} = \mathbf{D}\mathbf{A} + \mathbf{E} \ , \tag{4.3.1.3}$$

where $\mathbf{Y} = (Y^{(1)}, \cdots, Y^{(p)})$, and $Y^{(j)}$ represents K independent observations on variable j. Here $\mathbf{A} = (\alpha^{(1)}, \cdots, \alpha^{(p)})$ and $\alpha^{(j)} = (\mu_Y^{(j)}, \beta_1^{(j)}, \cdots, \beta_q^{(j)})'$, so that $\mathbf{A}$ is the matrix of regression (control) coefficients. Moreover, $\mathbf{D}$ is as in equation (2.1.35), K is the number of replications, and $\mathbf{E} = (\epsilon^{(1)}, \cdots, \epsilon^{(p)})$ where $\epsilon^{(j)}$ is the column vector of K residuals for the j^{th} variable so that each row of $\mathbf{E}$ is $\sim$ IID $N_p(\mathbf{0}, \Sigma)$ By similiar arguments (Seber (1984)), we estimate

$$\hat{\mathbf{A}} = (\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'\mathbf{Y} \ . \tag{4.3.1.4}$$

$$\hat{\Sigma} = \frac{(\mathbf{Y} - \mathbf{D}\hat{\mathbf{A}})'(\mathbf{Y} - \mathbf{D}\hat{\mathbf{A}})}{K - Q - 1} \ .$$

For both the models given in equations (4.3.1.2) and (4.3.1.3), we can see that the bulk of the computation involved in estimating model parameters lies in the inversion of $\mathbf{D}'\mathbf{D}$. In the next section we discuss efficient Gaussian elimination methods to accomplish this end.

4.3.1 Matrix Methods

In this section we discuss methods used to invert matrices of the form $\mathbf{D}'\mathbf{D}$, where $\mathbf{D}$ is $K \times Q$ and of column rank Q. These methods are based on *elementary row operations.* Such methods are called *Gaussian elimination methods.* We will discuss the Gauss-Jordan method, the *sweep* operator, symmetric sweeping, and Gaussian elimination.

Following Kennedy and Gentle (1980), we enumerate the following elementary row operations on a matrix:

1. Interchange two rows.

2. Multiply any row by a constant.

3. Add one row to another. To effect an elementary row operation, one can perform the operation on the identity matrix and simply premultiply this new matrix on a target matrix. These altered identity matrices are often called elementary matrices. Methods based on these elementary transformations are called Gaussian elimination methods.

Chvatal (1983) presents a clear exposition of Gaussian elimination methods. He condenses the elementary operations needed to invert a matrix into two basic matrices (a) the permutation matrix, and (b) the so-called *eta* matrix. The permutation matrix is an elementary matrix that simply interchanges two rows. The *eta* matrix is a succession of elementary matrices that zero selected elements of a matrix.

The Gauss-Jordan method of matrix inversion is based on the following observation. Let the matrix $\mathbf{T}$ be the product of successive multiplications of

those elementary matrices which must be applied to a square matrix $\mathbf{D'D}$ (where $\mathbf{D}$ is $K \times Q$ and of column rank Q) in order to reduce it to the identity matrix. That is, if $\mathbf{E}_j$ is the j^{th} elementary matrix, then, if

$$\mathbf{T} = \mathbf{E}_p \ldots \mathbf{E}_1$$

we have

$$\mathbf{T}(\mathbf{D'D}) = \mathbf{I}_Q$$

and

$$\mathbf{T}(\mathbf{D'D})(\mathbf{D'D})^{-1} = \mathbf{I}_Q (\mathbf{D'D})^{-1}$$

or

$$\mathbf{TI}_Q = (\mathbf{D'D})^{-1}$$

Now to implement this method, we merely augment $\mathbf{D'D}$ with $\mathbf{I}_Q$ and apply $\mathbf{T}$ to both. In the notation of Kennedy and Gentle (1980),

$$[(\mathbf{D'D}) \mid \mathbf{I}_Q] \underset{T}{\rightarrow} [\mathbf{I}_Q \mid (\mathbf{D'D})^{-1}]$$

Beaton (1964) exploited the fact that for each column reduction of $\mathbf{D'D}$ all the columns of the identity matrix remain in the transformed (augmented)

matrix. Essentially, he simply replaces the newly created identity column with its counterpart in the right hand side of the augmented matrix. This storage innovation allows for the inversion of a matrix in its own space. The operations which accomplish this are called *sweeps*. Beaton (see Seber (1977), pp. 351) offers the sweep operators. A matrix **G** is said to be swept on the k^{th} row and column if it has been transformed into a matrix $\mathbf{G}^{*} = ||g_{ij}^{'}||$ such that

$$g_{kk}^{'} = \frac{1}{g_{kk}}$$

$$g_{ik}^{'} = \frac{-g_{ik}}{g_{kk}} \qquad (i \neq k)$$

$$g_{kj}^{'} = \frac{g_{kj}}{g_{kk}} \qquad (j \neq k)$$

$$g_{ij}^{'} = g_{ij} - \frac{g_{ik} g_{kj}}{g_{kk}} \qquad (i, j \neq k)$$

Schatzoff et al. (1968), Seber (1977), and Kennedy and Gentle (1980) discuss the properties of the sweep operator in varying detail. A useful result, given in Kennedy and Gentle is the following. Let **G** be the augmented matrix

$$\mathbf{G} = \begin{bmatrix} \mathbf{D'D} & \mathbf{D'Y} \\ \mathbf{Y'D} & \mathbf{Y'Y} \end{bmatrix} , \qquad (4.3.1.5)$$

where **D** is $K \times Q$ (with column rank Q) and **Y** is $K \times p$. Now, if **G** is swept on the first Q rows and columns, then the resultant matrix $\mathbf{G}^{'}$ is

$$\mathbf{G}' = \begin{bmatrix} (\mathbf{D}'\mathbf{D})^{-1} & (\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'\mathbf{Y} \\ -\mathbf{Y}'\mathbf{D}(\mathbf{D}'\mathbf{D})^{-1} & \mathbf{Y}'\mathbf{Y} - \mathbf{Y}'\mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'\mathbf{Y} \end{bmatrix} . \qquad (4.3.1.6)$$

In particular, if $\mathbf{D}$ and $\mathbf{Y}$ are "centered" (that is, if the sample mean of each variable being subtracted from each obsevation), then

$$\mathbf{G}' = \begin{bmatrix} (\mathbf{D}'\mathbf{D})^{-1} & \hat{\mathbf{A}}' \\ -\hat{\mathbf{A}} & RSS \end{bmatrix} ,$$

where $\hat{\mathbf{A}}$ is as given in equation (4.3.1.4) and RSS is the matrix of residual sums and cross products.

The matrix $\mathbf{G}$, is symmetric. The symmetry of $\mathbf{G}$ can be exploited by working only on the upper triangle. This method is called the *symmetric sweep*. It is due to Steifel (1963) and for each pivot on matrix $\mathbf{G}$, we get the matrix $\mathbf{G}'$ such that

$$g_{kk}' = \frac{-1}{g_{kk}}$$

$$g_{ik}' = g_{ki}' = g_{ik}g_{kk}' \quad (i \neq k)$$

$$g_{ij}' = g_{ji}' = g_{ij} + g_{ik}g_{kj}' \quad (i,j \neq k)$$

Sweep and symmetric sweep operators produce both the regression (or control) coefficients and the RSS matrix. For some applications, such as our problem, there is no need to calculate all the control coefficients for every sweep. In these applications, the primary interest is the RSS matrix. The desired result can be obtained by applying Gaussian elimination to the matrix $\mathbf{G}$, given in equation (4.3.1.5). Seber (1977, page 304) shows the

desired result for the case when Y is $1 \times K$. The result for a $K \times p$ matrix $\mathbf{Y}$ is suggested by Kennedy and Gentle (1980, equation (7.21)) and is reproduced here as equation (4.3.1.6). To prove that Gaussian elimination produces the appropriate RSS matrix, one premultiplys the matrix $\mathbf{G}$ of (4.3.1.5) by the matrix

$$\begin{bmatrix} \mathbf{T} & \mathbf{0} \\ -\mathbf{Y}'\mathbf{D}(\mathbf{D}'\mathbf{D})^{-1} & \mathbf{I}_p \end{bmatrix} ,$$

where $\mathbf{T}$ is the lower triangular, nonsingular matrix that reduces $\mathbf{D}'\mathbf{D}$ to unit lower diagonal. If $\mathbf{D}'\mathbf{D}$ is nonsingular then the existence of $\mathbf{T}$ is assured, see Seber (1977, Chapter 11).

4.3.2. Generation of All Possible Regressions

In order to implement the criteria obtained in Chapter 3, a methodology is required to generate the needed regression information from each subset of regressors. Several systematic procedures have been suggested. Garside (1971), Schatzoff et al. (1968), and Furnival (1971) offer methods based on a binary coding for each subset of regressors. Furnival and Wilson (1974) offer algorithms which produce all possible regressions in orders they call natural, lexicographic, and familial. These orderings are based on the systematic search of a structure called a regression tree.

As we have seen in the previous section, Gaussian elimination and/or sweeps (henceforth, we refer to individual eliminations as "pivots") can produce regression information for some subset of the regressors (controls). We need to generate a sequence of pivots on carefully stored matrices, such that every possible subset of controls is considered. Following Furnival and

Wilson (1974), we form a regression tree as follows (see Figure 3). At the root of the tree is the original covariance matrix. At this point no variables (regressors) have been allowed to pivot into the model. The notation 123 signifies that there are 3 candidate controls. Dark lines on the tree signify pivots. For instance, the dark line emanating from the root signifies a pivot on variable 1. Note that the resultant notation, after this pivot, is 23.1. Integers after the dot signify those variables included in the model; the variables appearing before the dot are not yet in the model. Dotted lines represent the deletion of a variable from the model. For instance, the dotted line emanating from the root signifies the deletion of variable 1. Note that this deletion results in the model 23. At each node, we either pivot a variable into the model or delete a variable from the model. These pivots and deletions are carried out until all possible subsets of models have been considered. Furnival and Wilson point out that this tree can be traversed in any "biologically" feasible order. That is, we require only that a father be born before a son. If we search the tree horizontally from level to level, this is the so-called natural order. Storage savings are possible if if we use the so called lexicographic (dictionary-like) order. Furnival and Wilson also offer a (a) binary ordering which amounts to a counting process in base two, and a (b)familial ordering in which both horizontal and vertical elements are combined. Table 4.4 shows the order of the regressions produced by the four methods addressed above. Furnival and Wilson offer algorithms for each ordering.

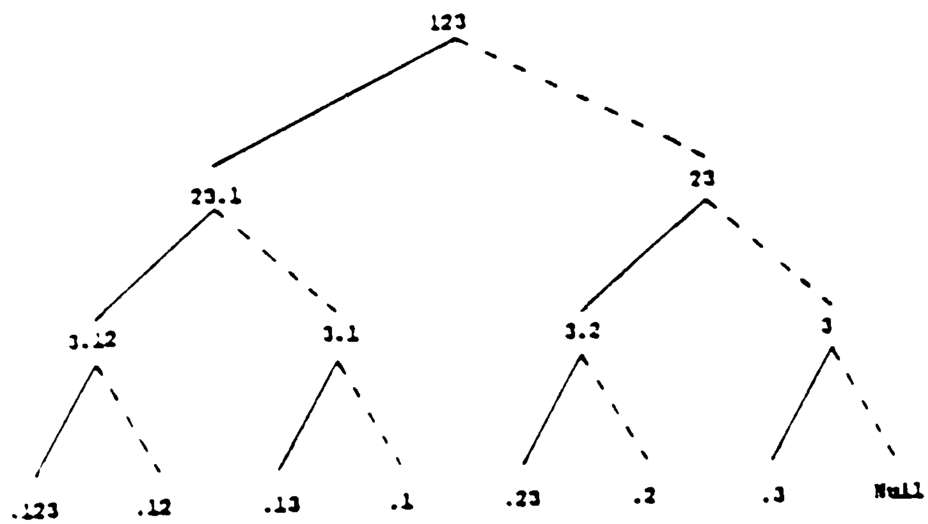


Figure 3. Regression Tree

Table 4.4 Sequences of Regressions

Natural	Lexicographic	Binary	Familial
1	1	1	1
2	12	2	2
3	123	12	3
4	1234	3	4
12	124	13	12
13	13	23	13
14	134	123	23
23	14	4	123
24	2	14	14
34	23	24	24
123	234	124	24
124	24	34	124
134	3	134	134
234	34	234	234
1234	4	1234	1234

4.3.3. Multivariate Generalization of All Possible Regressions

We employed a binary search of the regression tree. This search is a direct implemetation of the binary algorithm given by Furnival and Wilson (1974, page 504). The input matrix was the sample covariance matrix of controls and responses. At each pivot we calculated the generalized residual variance and computed the value of the selection criterion. We employed a variant of Gaussian elimination (also found in Furnival and Wilson) that exploits the symmetry of the covariance matrix by operating only on the upper triangle. We chose Gaussian elimination over sweeping to save on computation of the regression coefficients. These coefficients are only needed for the optimal subset.

CHAPTER 5
EXPERIMENTAL RESULTS

In this chapter we discuss the experimental portion of the research. We summarize the results of our simulation experiments and conclude the chapter by discussing a few experimental excursions designed to examine the underlying assumptions of our analysis.

5.1 Summary of Experimental Results

As mentioned in Section 4.1, we experimented with two different types of closed queueing networks. For purposes of discussion, we refer to the first type of system (no blocking at the CPU) as a type I network and the second type (blocking at the CPU) as a type II network. Models 1 and 2 (models 4 and 5 of Lavenberg, Moeller, and Welch (1978)) were type I models, while models 3 and 4 (models 15 and 16 of Lavenberg et al.) were type II models. All models were discrete-event simulations written in SLAM (Pritsker 1986) and a FORTRAN listing of both types is given in Appendix 5.

One basic aim of this research is to provide a simulation practitioner with a methodology whereby he can select the "best" subset of controls to

use to in constructing a confidence region for the steady-state mean vector of responses. The analysis program we developed to accomplish this end is given in Appendix 7.

For type I networks, it is possible to obtain analytically ,the steady-state expected values of the response vector $\left(\mathrm{R}(t),\ \mathrm{U}_{\mathrm{CPU}}(t)\right)$. We used the software package CAN-Q (Solberg (1980)) to calculate these values. Since it is impossible to run our simulations to infinity, we settled for long run lengths. For type I models, we chose a run length of 20,000 time units. For type II models, we chose a run length of 30,000 time units. We started the collection of statistics after 2,000 time units in an effort to minimize the effects of initialization bias. We chose the longer run length for type II models because of the presence of blocking at the CPU. When steady-state expected values (μ_Y) were not available, we used the grand mean vector of the 2,000 replications ($\overline{Y}(2{,}000)$) as the population mean. We report our results relative to both μ_Y and ($\overline{Y}(2{,}000)$). Table 5.1 summarizes this information for all four models.

Table 5.1 Mean Responses for the Queueing Systems Used in the Experimental Evaluation

	Steady-State Mean		$\bar{Y}(2{,}000)$	
Model	R(t)	$U_{CPU}(t)$	R(t)	$U_{CPU}(t)$
1	36.13	.918	36.04	.9177
2	81.71	.413	81.14	.4128
3	*	*	247.06	.3590
4	*	*	85.92	.6625

The candidate controls chosen for the type I networks were the four standardized work variables (collected at all stations) and the three standardized routing variables. Note that the analysis program includes a tolerance check on incoming variables. This check precludes multicollinearity problems which would result if all three routing variables were included in the model. We chose only seven of the available control variables as candidates for the type II models. This was done to keep the analysis at a comparable dimensionality across both types of networks. We chose as candidates the standardized work variables for the CPU and the two busiest disk drives; also we used four standardized routing variables (excluding only routing to less frequented disk drives).

For each model we ran 2,000 independent replications. The first 1000 replications were used to yield 50 meta-experiments of replication size 20. All 2,000 replications were used to produce 50 metaexperiments of replication size 40. We report (with respect to both μ_Y and $\bar{Y}(2{,}000)$) estimated coverage probabilities and estimated volume reductions for both estimators ($\bar{Y}(\hat{\beta})$ and $\bar{Y}(\hat{\gamma})$). Nominal coverage was 90%. Tables 5.2 and 5.3 summarize our experiments.

Table 5.2 Performance of the Controlled Point and Confidence Region Estimators for K=20 Replications of the Selected Queueing Systems

	Coverage Probability (%)				Volume	
	Steady-state mean		$\bar{Y}(2{,}000)$		Reduction (%)	
Model	$\bar{Y}(\hat{\beta})$	$\bar{Y}(\hat{\gamma})$	$\bar{Y}(\hat{\beta})$	$\bar{Y}(\hat{\gamma})$	$\bar{Y}(\hat{\beta})$	$\bar{Y}(\hat{\gamma})$
1	78	80	86	86	73	45
2	28	64	78	80	83	52
3	*	*	82	90	61	43
4	*	*	83	84	46	34

Table 5.3 Performance of the Controlled Point and Confidence Region Estimators for K=40 Replications of the Selected Queueing Systems

	Coverage Probability (%)				Volume	
	Steady-state mean		$\bar{Y}(2{,}000)$		Reduction (%)	
Model	$\bar{Y}(\hat{\beta})$	$\bar{Y}(\hat{\gamma})$	$\bar{Y}(\hat{\beta})$	$\bar{Y}(\hat{\gamma})$	$\bar{Y}(\hat{\beta})$	$\bar{Y}(\hat{\gamma})$
1	58	76	84	88	76	61
2	0	36	80	94	86	69
3	*	*	84	90	63	53
4	*	*	88	86	47	41

We observe that in all cases $\bar{Y}(\hat{\gamma})$ covered the steady-state mean better than $\bar{Y}(\hat{\beta})$. Further, $\bar{Y}(\hat{\gamma})$ offered comparable if not superior, coverage relative to $\bar{Y}(\hat{\beta})$ when $\bar{Y}(2{,}000)$ is taken as the target mean. This improvement in reliability is probably due to the conservatism of $\bar{Y}(\hat{\gamma})$ to $\bar{Y}(\hat{\beta})$ as reflected in the realized volume reductions.

5.2 Examination of the Assumptions Underlying the Application of Control Variables

Two major assumptions come into play as one applies the estimators $\bar{Y}(\hat{\beta})$ and $\bar{Y}(\hat{\gamma})$. Confidence interval procedures for both estimators are based on the assumption of joint normality between the responses and the controls. The confidence interval procedure based on $\bar{Y}(\hat{\gamma})$ assumes that the mean vector and dispersion matrix of controls is known. In the case of the controls we applied, these quantities are known only asymptotically. We also assumed that the runs were of sufficient length that the response vector could be assumed to be in steady-state.

We carried out three excursions from the primary analysis to gauge the effects of the underlying assumptions. We hoped these experiments would shed some light on the sensitivity of the procedures to the underlying assumptions. First, we looked at the situation where the responses and controls were distributed as jointly normal random variables with all means and covariances known exactly. We calculated the sample covariance matrix of responses and controls for 2,000 replications of model 2 (chosen arbitrarily). We took this covariance matrix to be the population covariance matrix of responses and controls. Next, we generated 2,000 independent, normally distributed random vectors based on this structure. We repeated our basic analysis for replication sizes of 20 and 40. In this experiment the means of the responses as well as the controls are known exactly. The results are Table 5.4

Table 5.4 Performance of the Controlled Point and Confidence Region Estimators when Multivariate Normality is Ensured

	Coverage Probability (%)		Volume Reduction (%)	
Replications	$\bar{Y}(\hat{\beta})$	$\bar{Y}(\hat{\gamma})$	$\bar{Y}(\hat{\beta})$	$\bar{Y}(\hat{\gamma})$
20	89	93	81	50
40	92	94	85	69

We observe that under ideal conditions both estimators deliver nominal coverage and $\bar{Y}(\hat{\gamma})$ is more conservative. We feel that the results above validate $\bar{Y}(\hat{\gamma})$ as a viable estimator.

Next, we wanted to see if we could determine whether it was the lack of normality in the responses or insufficient run lengths which degraded estimator perfromance in the actual simulations. To accomplish this end, we applied normalizing transformations (Box, Hunter, and Hunter (1978), and (Anderson and Mclean (1974)) in an effort to make the data appear more normally distributed. We took the natural logarithm of the response times and applied the transformation $z =$ arcsin $(\sqrt{U_{CPU}(t)})$ to the CPU utilizations. The results are Table 5.5. Coverage is relative to the grand average of the 2,000 transformed response vectors $(\bar{Y}_z(2{,}000))$.

Table 5.5 Performance of the Controlled Point and Confidence Region Estimators Under Normalizing Transformations of Queueing Simulation Responses

	Coverage Probability (%) $\bar{Y}_z(2{,}000)$		Volume Reduction (%)	
Replications	$\bar{Y}(\hat{\beta})$	$\bar{Y}(\hat{\gamma})$	$\bar{Y}(\hat{\beta})$	$\bar{Y}(\hat{\gamma})$
20	84	85	82	51
40	74	90	86	70

Comparing Tables 5.3 and 5.4 to Table 5.5, we see that there seems to be an indication that $\bar{Y}(\hat{\gamma})$ is less sensitive to departures from normality than $\bar{Y}(\hat{\beta})$.

Finally, we were interested in the effects of run length. We contrast the performance measures at run lengths of 5,000 and 20,000 time units, respectively. The results are Table 5.6. The replication level is 20.

Table 5.6 Performance of the Controlled Point and Confidence Region Estimators for Queueing Simulations of Different Run Lengths

	Coverage Probability (%)				Volume Reduction (%)	
	Steady-state mean		$\bar{Y}(2{,}000)$			
Run Length	$\bar{Y}(\hat{\beta})$	$\bar{Y}(\hat{\gamma})$	$\bar{Y}(\hat{\beta})$	$\bar{Y}(\hat{\gamma})$	$\bar{Y}(\hat{\beta})$	$\bar{Y}(\hat{\gamma})$
5000	18	33	90	92	57	39
30000	28	64	78	80	83	52

We observe that the increase in run length yields improvements in the coverage of the steady-state mean vector for both estimators. We also observe a degradation in coverage about $\bar{Y}(2{,}000)$ when the run length is increased. Volume reductions appear to be significantly larger for the long runs.

CHAPTER 6
CONCLUSIONS AND RECOMMENDATIONS

6.1 Overview

This research offers a solution to the general problem of optimal selection of control variates. We offer solutions for two different cases of the general problem: (a) when the covariance matrix of the controls is unknown, and (b) when the covariance matrix of the controls is known and is incorporated into point and confidence region estimators. For the second case we introduce a new estimator that we represent by the symbol $\bar{Y}(\hat{\gamma})$. Under the assumption that the responses and the controls are jointly normal, we have established the unbiasness of this new estimator, and we have derived its dispersion matrix. We have implemented a selection algorithm which locates the optimal subset of controls. The algorithm is based on criteria we derive for the two cases listed above. We have introduced a new class of controls which we call "routing variables". We derived the asymptotic distribution of these controls as well as their asymptotic mean and variance. Finally, we have investigated the performance of the selection algorithm and we have contrasted the estimators $\bar{Y}(\hat{\beta})$ and $\bar{Y}(\hat{\gamma})$.

6.2 Conclusions

We conclude that the selection algorithm delivers a workable set of controls that yield minimum expected squared volume. Under ideal conditions, confidence region procedures based on both estimators deliver nominal coverage and significant volume reduction. The estimator $\overline{Y}(\hat{\gamma})$ appears to be more conservative. The conservatism of $\overline{Y}(\hat{\gamma})$ greatly enhances its reliability when the underlying assumptions are violated.

Routing variables prove to be a significant new class of controls variables, in that they entered every model fitted during the course of the analysis. They are easy to implement and should be considered as candidate controls in any simulation that contains probabilistic branching.

6.3 Recommendations

There are several avenues of potential future research that present themselves.

1. In the case where the covariance matrix of the controls is known, we offer an unbiased estimate of the dispersion matrix of $\overline{Y}(\hat{\gamma})$ which we call $\hat{\hat{\Sigma}}$. The theoretical properties of $\hat{\hat{\Sigma}}$ have yet to be established. It would also be of interest to investigate other estimators of $\tilde{\tilde{\Sigma}}$.

2. We employed a "all-regressions" approach to find the optimal subset of controls. Generalization of a search scheme that avoids total enumeration of the subsets would be useful. A multivariate generalization of the branch and bound algorithm suggested by Furnival and Wilson (1974) is immediately suggested.

3. Dr. Bruce Schmeiser has suggested that the response variables be scaled prior to analysis to reflect the preferences of whoever uses the model to make a decision. Research aimed at incorporating the preference structure and risk adversion of the decision maker would be of interest.

BIBLIOGRAPHY

BIBLIOGRAPHY

Aitkin, M. A., "Simultaneous Inference and the Choice of Variable Subsets in Multiple Regression," *Technometrics*, vol. 16, pp. 221-227, 1974.

Akaike, H., "Information Theory and an Extension of the Maximum Likelihood Principle," *2nd International Symposium on Information Theory*, pp. 267-281, Akademiai Kiado, Budapest, 1973.

Allen, D. M., "Mean Square Error of Prediction as a Criterion for Selecting Variables," *Technometrics*, vol. 13, pp. 469-475, 1971.

Anderson, T. W., *An Introduction to Multivariate Statistical Analysis,* John Wiley, New York, New York, 1984.

Arvensen, J. N., "Jackknifing U-Statistics," *Annals of Mathematical Statistics*, vol. 40, pp. 2076-2100, 1969.

Bauer, Kenneth W., A Monte Carlo Study of Dimensionality Assessment and Factor Interpretation in Principal Components Analysis, Unpublished Masters Thesis, Air Force Institute of Technology, 1981.

Beaton, A. E., The Use of Special Matrix Operators in Statistical Calculus. Research Bulletin RB-64-51, Educational Testing Service, Princeton, New Jersey, 1964.

Box, George E., William G. Hunter, and J. Stuart, *Statistics for Experimenters,* John Wiley and Sons, New York, New York, 1978.

Cheng, R. C. H., "Analysis of Simulation Experiments under Normality Assumptions," *Journal of the Operational Research Society*, vol. 29, pp. 493-497, 1978.

Chvatal, Vasek, *Linear Programming,* W.H. Freeman and Company, New York, New York, 1980.

Crane, M. and A. J. Lemoine, "An Introduction to the Regenerative Method for Simulation Analysis," in *Lecture Notes in Control and Information Sciences*, Springer-Verlag, Berlin, Germany, 1977.

Draper, N. and H. Smith, *Applied Regression Analysis, Second Edition,* John Wiley and Sons, New York, New York, 1981.

Eakle, J. D., Regenerative Analysis Using Internal Controls, Unpublished Ph.D. Dissertation, Mechanical Engineering Department, University of Texas, Austin, Texas, 1982.

Flury, Bernard and Hans Reidwyl, "T ** 2 Tests and the Linear Two-Group Discriminant Function, and their Computation by Linear Regression," *The American Statistician*, vol. 39, pp. 20-25, February, 1985.

Furnival, G. M. and R. W. Wilson, "Regression by Leaps and Bounds," *Technometrics*, no. 16, pp. 499-511, 1974.

Furnival, George M., "All Possible Regressions with Less Computation," *Technometrics*, vol. 13, pp. 403-408, May 1971.

Gabriel, K. R., "Simultaneous Test Procedures - Some Theory of Multiple Comparisons," *Annals of Mathematical Statistics*, vol. 40, pp. 224-250, 1969.

Hocking, R. R., "The Analysis and Selection of Variables in Linear Regression," *Biometrics*, vol. 32, pp. 1-49, March, 1976.

Hocking, R. R., "Developments in Linear Regression Methodology: 1959-1982," *Technometrics*, vol. 25, pp. 219-230, August, 1983.

Hoerl, A. E. and R. W. Kennard, "Ridge Regression: Biased estimation for non-orthogonal problems," *Technometrics*, vol. 12, pp. 55-68, 1970.

Hogg, R. V. and A. T. Craig, *Introduction to Mathematical Statistics,* Macmillan Company, London, England, 1970.

Iglehart, D. L., "The Rengerative Method for Simulation Analysis," in *Current Trends in Programming Methodology, Vol III* , ed. K. M. Chandy and R. Yeh , Prentice Hall, Englewood Cliffs, New Jersey, 1978.

Iglehart, D. L. and P. A. W. Lewis, "Regenerative Simulation with Internal Controls," *Journal of the Association of Computing Machinery*, vol. 26, no. 2, pp. 271-282, 1979.

Johnson, R. A. and D. W. Wichern, *Applied Multivariate Analysis,* Prentice Hall, Inc. , Englewood Cliffs, New Jersey, 1982.

Jollife, I. T., "Disgarding Variables in Principal Components Analysis: Part I . Artifical Data," *Applied Statistics*, vol. 21, pp. 160-173, 1972.

Kennedy, W. J. and J. E. Gentle. *Statistical Computing,* Marcel Dekker, Inc., New York, New York, 1980.

Kenney, J. F. and E. S. Keeping, *Mathematics of Statistics Part II,* Van Nostrand Co. Inc., New York, New York, 1951.

Kleijnen, J. P. C., *Statistical Techniques in Simulation, Part 1 and 2,* Marcel Deckker, New York, New York, 1975.

Lavenberg, S. S., T. L. Moeller, and P. D. Welch, Statistical Results on Multiple Control Variables with Application to Variance Reduction in Queueing Network Simulation, IBM Research Report RC-7423, Yorktown Heights, New York, 1978.

Lavenberg, S. S., T. L. Moeller, and C. H. Sauer, "Concomitant Control Variables Applied to the Regenerative Simulation of Queueing Systems," *Operations Research*, vol. 27, no. 1, pp. 134-160, 1979.

Lavenberg, S. S. and P. D. Welch, "A Perpective on the Use of Control Variables to Increase the Efficiency of Monte Carlo Simulations," *Management Sciences*, vol. 27, pp. 322-334, March 1981.

Lavenberg, S. S., T. L. Moeller, and P. D. Welch, "Statistical Results on Control Variables with Application to Queueing Network Simulation," *Operations Research*, vol. 30, pp. 182-202, Jan-Feb 1982.

Lavenberg, Stephen S., T. L. Moeller, and P. D. Welch, "Statistical Results on Control Variables with Applications to Queueing Network Simulation," *Operation Research*, vol. 30, pp. 182-202.

Lindley, D. V., "The Choice of Variables in Multiple Regression," *Journal of the Royal Statistical Society*, vol. B30, pp. 31-53, 1968.

Mallows, C. L., "Some Comments on C sub p," *Technometrics*, vol. 15, pp. 661-675, 1973.

Mantel, N., "Why Stepdown Procedures in Variable Selection," *Technometrics*, vol. 12, pp. 621-625, 1970.

McCabe, George P., "Evaluation of Regression Coefficient Estimates Using alpha-acceptabilty," *Technometrics*, vol. 20, pp. 131-139, May 1978.

McCabe, George P., "Principal Variables," *Technometrics*, vol. 26, pp. 137-144, May, 1984.

McKay, R. J., "Variable Selection in Multivariate Regression: An Application of Simultaneous Test Procedures," *Journal of the Royal Statistical Society*, vol. B 39, pp. 371-380, 1977.

Miller, Rupert G., "The jackknife - a review," *Biometrika*, vol. 61, pp. 1-15, 1974.

Muirhead, R. J., *Aspects of Multivariate Statistical Theory,* John Wiley and Sons, New York, New York, 1982.

Neter, J., W. Wasserman, and M. H. Knuter, *Applied Linear Regression Models,* Richard D. Erwin, Inc., Homewood, Illinois, 1983.

Neuts, Marcel F., *Probability,* Allyn and Bacon, Inc., Boston, Mass., 1973.

Nova, A. M. Porta, A Generalized Approach to Variance Reduction in Discrete-event Simulation using Control Variables, Unpublished Ph.D. Dissertation, Department of Mechanical Engineering, The University of Texas, Austin, Texas, 1985.

Nozari, A., S. F. Arnold, and C. D. Pegden, "Control Variates for Multipopulation Experiments," *IIE Transactions*, vol. 16, pp. 159-169, June, 1984.

Pritsker, A. Alan B., *Introduction to Simulation and SLAM,* Halsted Press, New York, New York, 1986.

Rao, C. R., "Least Squares Theory using an Estimated Dispersion Matrix and its Application to Measurement of Signals," *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probabilty*, vol. I, pp. 355-372, University of California Press, Berkeley, California, 1967.

Rubinstein, Reuven Y. and Ruth Marcus, "Efficiency of Multivariate Control Variates in Monte Carlo Simulation," *Operations Research*, vol. 33, pp. 661-677, May-June 1985.

Schatzoff, M., S. Fienberg, and R. Tsao, "Efficient Calculations of All-possible Regressions," *Technometrics*, vol. 10, pp. 768-779, 1968.

Schwarz, G., "Estimating the Dimension of a Model," *Annals of Statistics*, vol. 6, pp. 461-464, 1978.

Seber, G. A. F., in *Linear Regression Analysis*, John Wiley and Sons, New York, New York, 1977.

Seber, G. A. F., *Multivariate Observations,* John Wiley and Sons, New York, New York, 1984.

Siotani, M., T. Hayakawa, and Y. Fujikoshi, *Modern Multivariate Statistical Analysis: A Graduate Course and Handbook,* American Sciences Press, Columbus, Ohio, 1985.

Solberg, James J., CAN-Q User's Manual, School of Industrial Engineering, Purdue University West Lafayette, Indiana, 1980.

Thompson, Mary L., "Selection of Variables in Multiple Regression: A Review and Evaluation," *International Statistical Review,* vol. 46, pp. 1-19, 1978.

Venkatraman, Sekhar, Application of the Control Variate Technique to multiple Simulation Output Analysis, Department of Mechanical Engineering, The University of Texas, Austin, Texas, 1983.

Venkatraman, Sekhar and James R. Wilson, "The Efficiency of Control Variates in Multiresponse Simulation," *O.R. Letters*, vol. 5, no. 1, pp. 37-42, 1986.

Webster, J. T., R. F. Gunst, and R. L. Mason, "Latent Root Regression Analysis," *Technometrics*, vol. 16, pp. 513-522, 1974.

Welch, Peter D., "The Statistical Analysis of Simulation Results," in *Computer Preformance Modeling Handbook*, Academic Press Inc., New York, New York, 1983.

Wilson, J. R. and A. A. B. Pritsker, "Variance Reduction in Queueing Simulation using Generalized Concomitant Variables," *Journal of Statistical Computation and Simulation*, vol. 19, pp. 129-153, 1984a.

Wilson, J. R. and A. A. B. Pritsker, "Experimental Evaluation of Variance Reduction Techniques for Queueing Simulation using Generalized Concominant Variables," *Management Science*, vol. 30, pp. 1459-1472, Dec 1984b.

Wilson, James R., Variance Reduction Techiques for the Simulation of Queueing Networks, Technical Report , Mechanical Engineering Department, University of Texas, Austin, Texas, 1982.

Wilson, James R., "Variance Reduction Techniques for Digital Simulation," *American Journal of Mathematical and Management Sciences*, vol. 1, pp. 227-312, 1984 .

APPENDICES

Appendix 1: Derivation of Equation (3.2.1.10)

Now, define $\mathbf{Z} = \text{vec } \mathbf{X} = \text{vec } (X_1, \cdots, X_K)$ so that $\mathbf{Z}$ is the (KQ) dimensional column vector formed by stacking the K X_j one upon another. Now, the condition $\left\{ X_j = x_j \right\}$ can be more compactly expressed as $\mathbf{Z} = z$. Also, define $\mu_Z = \text{vec } (\mathbf{1}_K \otimes \mu_X)$. Write

$$\bar{\mathbf{Y}}(\hat{\cdot}) = \left[Y_1, \cdots, Y_K \right] \tilde{\mathbf{H}} = \mathbf{Y}\tilde{\mathbf{H}} \tag{A.1.1}$$

and

$$\text{vec } \bar{\mathbf{Y}}(\hat{\cdot}) = \text{vec } (\mathbf{Y}\tilde{\mathbf{H}}) = (\tilde{\mathbf{H}}' \otimes I_p) \text{ vec } \mathbf{Y} \;. \tag{A.1.2}$$

Now

$$\text{cov}[\text{ vec } \mathbf{Y}\tilde{\mathbf{H}}] = E_X \left[\text{cov} \left[\text{ vec } \mathbf{Y}\tilde{\mathbf{H}} \,|\, \mathbf{Z} = z \right] \right] + \text{cov} \left[E \left[\text{ vec } \mathbf{Y}\tilde{\mathbf{H}} \,|\, \mathbf{Z} = z \right] \right] \tag{A.1.3}$$

Examining the first term of the right hand side of equation (A.1.3)

$$E_X \left[\text{cov} \left[\text{ vec } \mathbf{Y}\tilde{\mathbf{H}} \,|\, \mathbf{Z} = z \right] \right] = E_X \left[\text{cov} \left[(\tilde{\mathbf{H}}' \otimes I_p) \text{ vec } \mathbf{Y} \,|\, \mathbf{Z} = z \right] \right] \tag{A.1.4}$$

$$= E_X \left[(\widetilde{\mathbf{H}}' \otimes I_p) \text{cov} \left[\text{ vec } \mathbf{Y} \,|\, \mathbf{Z} = z \right] (\widetilde{\mathbf{H}}' \otimes I_p)' \right] \tag{A.1.5}$$

$$= E_X \left[(\widetilde{\mathbf{H}}' \otimes I_p)(I_K \otimes \Sigma_{Y\,|\,X})(\widetilde{\mathbf{H}}' \otimes I_p)' \right] \tag{A.1.6}$$

$$= E_X \left[(\widetilde{\mathbf{H}}' \otimes \Sigma_{Y\,|\,X})(\widetilde{\mathbf{H}} \otimes I_p) \right] \tag{A.1.7}$$

$$= E_X \left[(\widetilde{\mathbf{H}}' \widetilde{\mathbf{H}} \otimes \Sigma_{Y\,|\,X}) \right] = \Sigma_{Y\,|\,X} E_X \left[\widetilde{\mathbf{H}}' \widetilde{\mathbf{H}} \right] \tag{A.1.8}$$

Now

$$\widetilde{\mathbf{H}}' \widetilde{\mathbf{H}} = \left[K^{-1} + (K-1)^{-1} (\overline{\mathbf{X}} - \mu_{\mathrm{X}})' \Sigma_{XX}^{-1} \mathbf{S}_{XX} \Sigma_{XX}^{-1} (\overline{\mathbf{X}} - \mu_{\mathrm{X}}) \right] \tag{A.1.9}$$

$$E \left[\widetilde{\mathbf{H}}' \widetilde{\mathbf{H}} \right] = K^{-1} + E \left[\text{tr} (K-1)^{-1} (\overline{\mathbf{X}} - \mu_{\mathrm{X}})' \Sigma_{XX}^{-1} \mathbf{S}_{XX} \Sigma_{XX}^{-1} (\overline{\mathbf{X}} - \mu_{\mathrm{X}}) \right] \tag{A.1.10}$$

by the properties of the trace (tr). Now

$$E \left[\text{tr} (K-1)^{-1} (\overline{\mathbf{X}} - \mu_{\mathrm{X}})' \Sigma_{XX}^{-1} \mathbf{S}_{XX} \Sigma_{XX}^{-1} (\overline{\mathbf{X}} - \mu_{\mathrm{X}}) \right] = \tag{A.1.11}$$

$$(K-1)^{-1}\mathrm{tr}\left[E\left[(\overline{\mathbf{X}}-\mu_{\mathrm{X}})'\Sigma_{XX}^{-1}\mathbf{S}_{XX}\Sigma_{XX}^{-1}(\overline{\mathbf{X}}-\mu_{\mathrm{X}})\right]\right] = \tag{A.1.12}$$

$$(K-1)^{-1}\mathrm{tr}\left[E\left[(\overline{\mathbf{X}}-\mu_{\mathrm{X}})(\overline{\mathbf{X}}-\mu_{\mathrm{X}})'\Sigma_{XX}^{-1}\mathbf{S}_{XX}\Sigma_{XX}^{-1}\right]\right] = \tag{A.1.13}$$

$$(K-1)^{-1}\mathrm{tr}\left[E\left[(\overline{\mathbf{X}}-\mu_{\mathrm{X}})(\overline{\mathbf{X}}-\mu_{\mathrm{X}})'\right]\Sigma_{XX}^{-1}E\left[S_{XX}\right]\Sigma_{XX}^{-1}\right] \tag{A.1.14}$$

$$(K(K-1))^{-1}\mathrm{tr}\left[\Sigma_{XX}\Sigma_{XX}^{-1}\Sigma_{XX}\Sigma_{XX}^{-1}\right] = \tag{A.1.15}$$

$$(K(K-1))^{-1}\mathrm{tr}\left[I_Q\right] = (K(K-1))^{-1}Q\ . \tag{A.1.16}$$

so algebra shows that (A.1.8) (hence (A.1.4)) becomes

$$\Sigma_{Y\,|\,X}E_X\left[\widetilde{\mathbf{H}}'\widetilde{\mathbf{H}}\right] = \frac{K+Q-1}{K(K-1)}\Sigma_{Y\,|\,X} \tag{A.1.17}$$

Next, we examine the second term of the right hand side of (A.1.3)

$$\mathrm{cov}\left[E\left[\text{ vec }\mathbf{Y}\widetilde{\mathbf{H}}\,|\,\mathbf{Z}=z\right]\right] = \mathrm{cov}\left[E\left[(\widetilde{\mathbf{H}}'\otimes I_p)\text{ vec }\mathbf{Y}\,|\,\mathbf{Z}=z\right]\right] = \tag{A.1.18}$$

Equation (3.2.18) allows

$$\operatorname{cov}\left[(\widetilde{\mathbf{H}}'\otimes I_p)E\left[\text{ vec }\mathbf{Y}\,|\,\mathbf{Z}=z\right]\right]= \tag{A.1.19}$$

$$\operatorname{cov}\left[(\widetilde{\mathbf{H}}'\otimes I_p)\left[(\mathbf{1}_K\otimes\mu_Y)+(I_K\otimes\Sigma_{YX}\Sigma_{XX}^{-1})(\mathbf{Z}-\mu_Z)\right]\right]. \tag{A.1.20}$$

Now, as a direct result of equation (3.2.1.9)

$$\operatorname{cov}\left[\text{ vec }(\Sigma_{YX}\Sigma_{XX}^{-1}\mathbf{X}\widetilde{\mathbf{H}})\right]= \tag{A.1.21}$$

$$\Sigma_{YX}\Sigma_{XX}^{-1}(\operatorname{cov}\left[\text{ vec }\mathbf{X}\widetilde{\mathbf{H}}\right])\Sigma_{XX}^{-1}\Sigma_{XY}\ . \tag{A.1.22}$$

Now

$$\operatorname{cov}(\text{ vec }\mathbf{X}\widetilde{\mathbf{H}})=\operatorname{cov}(\mathbf{X}\widetilde{\mathbf{H}})= \tag{A.1.23}$$

$$E(\mathbf{X}\widetilde{\mathbf{H}}(\mathbf{X}\widetilde{\mathbf{H}})')-E(\mathbf{X}\widetilde{\mathbf{H}})E(\mathbf{X}\widetilde{\mathbf{H}})'= \tag{A.1.24}$$

$$E(\mathbf{X}\widetilde{\mathbf{H}}\widetilde{\mathbf{H}}'\mathbf{X}')-E(\mathbf{X}\widetilde{\mathbf{H}})E(\mathbf{X}\widetilde{\mathbf{H}})'\ . \tag{A.1.25}$$

First, we examine the first term on the right hand side of equation (A.1.25). Now

$$(\tilde{\mathbf{H}}\tilde{\mathbf{H}}') = K^{-2}\mathbf{1}_K\mathbf{1}_K{}' - K^{-1}\mathbf{1}_K(\bar{\mathbf{X}}-\mu_{\mathbf{X}})'\Sigma_{XX}^{-1}\mathbf{G}'(K-1)^{-1} -$$

$$- (K-1)^{-1}\mathbf{G}\Sigma_{XX}^{-1}(\bar{\mathbf{X}}-\mu_{\mathbf{X}})K^{-1}\mathbf{1}_K{}' + (K-1)^{-2}\mathbf{G}\Sigma_{XX}^{-1}(\bar{\mathbf{X}}-\mu_{\mathbf{X}})(\bar{\mathbf{X}}-\mu_{\mathbf{X}})'\Sigma_{XX}^{-1}\mathbf{G}'$$

(A.1.26)

We note that ($\mathbf{G}$ is as given in equation (3.2.1.45)

$$\mathbf{XG} = (\mathbf{X} - \bar{\mathbf{X}}\mathbf{1}_K{}')\mathbf{G} = (K-1)S_{XX} \qquad \text{(A.1.27)}$$

Hence

$$\mathbf{X}(\tilde{\mathbf{H}}\tilde{\mathbf{H}}')\mathbf{X}' = K^{-2}\mathbf{X}\mathbf{1}_K\mathbf{1}_K{}'\mathbf{X}' - \bar{\mathbf{X}}(\bar{\mathbf{X}}-\mu_{\mathbf{X}})'\Sigma_{XX}^{-1}S_{XX}{}'$$

$$- S_{XX}\Sigma_{XX}^{-1}(\bar{\mathbf{X}}-\mu_{\mathbf{X}})\bar{\mathbf{X}}' + S_{XX}\Sigma_{XX}^{-1}(\bar{\mathbf{X}}-\mu_{\mathbf{X}})(\bar{\mathbf{X}}-\mu_{\mathbf{X}})'\Sigma_{XX}^{-1}S_{XX}{}'$$

(A.1.28)

Looking at the first term in the right hand side of equation (A.1.28) we have

$$E(K^{-2}\mathbf{X}\mathbf{1}_K\mathbf{1}_K{}'\mathbf{X}') = K^{-2}E(\bar{\mathbf{X}}\bar{\mathbf{X}}') = K^{-1}\Sigma_{XX} + \mu_X\mu_X{}' \qquad \text{(A.1.29)}$$

Examination of the second term in the right hand side of (A.1.28) reveals

$$E\left[\bar{\mathbf{X}}(\bar{\mathbf{X}}-\mu_{\mathbf{X}})'\Sigma_{XX}^{-1}S_{XX}{}'\right] = E\left[\bar{\mathbf{X}}(\bar{\mathbf{X}}-\mu_{\mathbf{X}})'\right]\Sigma_{XX}^{-1}E\left[S_{XX}{}'\right] = \qquad \text{(A.1.30)}$$

$$E\left[\overline{\mathbf{X}}\overline{\mathbf{X}}' - \overline{\mathbf{X}}\mu_X{}'\right]\Sigma_{XX}^{-1}\Sigma_{XX} = \tag{A.1.31}$$

$$E\left[\overline{\mathbf{X}}\overline{\mathbf{X}}' - \mu_X\mu_X{}'\right] = K^{-1}\Sigma_{XX} \ . \tag{A.1.33}$$

Now we examine the fourth term on the right hand side of (A.1.28), we observe as a consequence of equation (3.2.1.9) that

$$E\left[S_{XX}\Sigma_{XX}^{-1}(\overline{\mathbf{X}}-\mu_{\mathrm{X}})(\overline{\mathbf{X}}-\mu_{\mathrm{X}})'\Sigma_{XX}^{-1}S_{XX}{}'\right] = \mathrm{cov}\left[S_{XX}\Sigma_{XX}^{-1}(\overline{\mathbf{X}}-\mu_{\mathrm{X}})\right] = \tag{A.1.34}$$

$$E\left[\mathrm{cov}\left[S_{XX}\Sigma_{XX}^{-1}(\overline{\mathbf{X}}-\mu_{\mathrm{X}})\,|\,S_{XX} = s_{xx}\right]\right] + \mathrm{cov}\left[E\left[S_{XX}\Sigma_{XX}^{-1}(\overline{\mathbf{X}}-\mu_{\mathrm{X}})\,|\,S_{XX} = s_{xx}\right]\right] = \tag{A.1.35}$$

$$E\left[\mathrm{cov}\left[S_{XX}\Sigma_{XX}^{-1}(\overline{\mathbf{X}}-\mu_{\mathrm{X}})\,|\,S_{XX} = s_{xx}\right]\right] = E\left[S_{XX}\Sigma_{XX}^{-1}\mathrm{cov}\left[\overline{\mathbf{X}}|\,S_{XX} = s_{xx}\right]\Sigma_{XX}^{-1}S_{XX}\right] = \tag{A.1.36}$$

$$E\left[S_{XX}\Sigma_{XX}^{-1}K^{-1}\Sigma_{XX}\Sigma_{XX}^{-1}S_{XX}\right] = E\left[K^{-1}S_{XX}\Sigma_{XX}^{-1}S_{XX}\right] . \tag{A.1.37}$$

Let $\mathcal{W} \sim \mathcal{W}_{\mathrm{Q}}(\mathrm{K}-1\ ,\ \mathrm{I}_{\mathrm{Q}})$, where $\mathcal{W}_{\mathrm{Q}}(\mathrm{K}-1\ ,\ \mathrm{I}_{\mathrm{Q}})$ is a random matrix

distributed as Wishart with $K-1$ degrees of freedom and expected matrix I_Q. Now, $S_{XX} \sim \mathcal{W}_Q(K-1, K-1^{-1}\Sigma_{XX})$, so $S_{XX} \sim (K-1)^{-1}\Sigma_{XX}^{1/2}\mathcal{W}\Sigma_{XX}^{1/2}$. Now continuing from equation (A.1.37)

$$E\left[K^{-1}S_{XX}\Sigma_{XX}^{-1}S_{XX}\right] = E\left[K^{-1}(K-1)^{-2}(\Sigma_{XX}^{1/2}\mathcal{W}\Sigma_{XX}^{1/2})\Sigma_{XX}^{-1}(\Sigma_{XX}^{1/2}\mathcal{W}\Sigma_{XX}^{1/2})\right] = \tag{A.1.38}$$

$$E\left[K^{-1}(K-1)^{-2}\Sigma_{XX}^{1/2}\mathcal{W}\mathcal{W}\Sigma_{XX}^{1/2}\right] = \mathrm{K}^{-1}(\mathrm{K}-1)^{-2}\Sigma_{XX}^{1/2}\mathrm{E}\left[\mathcal{W}\mathcal{W}\right]\Sigma_{XX}^{1/2} \tag{A.1.39}$$

Now we compute $E\left[\mathcal{W}\mathcal{W}\right]$ with the following arguments. Let $Q = \mathcal{W}\mathcal{W}$, first consider the $E[Q]_{ij}$ when $i = j$.

$$E\left[Q_{ii}\right] = E\left[\sum_{v=1}^{Q} \mathcal{W}_{iv}\mathcal{W}_{iv}\right] = \tag{A.1.40}$$

$$\sum_{v=1}^{Q} E\left[\mathcal{W}_{iv}^2\right] = \tag{A.1.41}$$

Using equation (4), page 90 of Muirhead (1982) (and some algebra), we get

$$= \sum_{v=1}^{Q} var\left[\mathcal{W}_{iv}\right] + \mathrm{E}\left[\mathcal{W}_{ii}^2\right] = (\mathrm{K}-1)(\mathrm{K}+\mathrm{Q})\ . \tag{A.1.42}$$

Now, consider when $i \neq j$

$$E\left[Q_{ii}\right] = E\left[\sum_{v=1}^{Q} w_{iv}\, w_{jv}\right] = \tag{A.1.43}$$

$$E\left[w_{ii}\, w_{ji}\right] + \mathrm{E}\left[w_{ij}\, w_{jj}\right] + \sum_{\substack{v=1 \\ v \neq i,j}}^{Q} \mathrm{E}\left[w_{iv}\, w_{jv}\right] = \tag{A.1.44}$$

Once again we apply equation (4), page 90 of Muirhead (1982) (and some algebra) we get

$$\mathrm{cov}\left[w_{ii}\, w_{ji}\right] + \mathrm{cov}\left[w_{ij}\, w_{jj}\right] + \sum_{\substack{v=1 \\ v \neq i,j}}^{Q} \mathrm{cov}\left[w_{iv}\, w_{jv}\right] = 0\ . \tag{A.1.45}$$

So, we have

$$E\left[Q\right] = (K-1)(K+Q)I_Q \tag{A.1.46}$$

and so equation (A.1.34) (the fourth term on the right hand side of equation (A.1.28) becomes

$$E\left[K^{-1} S_{XX} \Sigma_{XX}^{-1} S_{XX}\right] = \frac{(K-1)(K+Q)}{(K-1)^2} \Sigma_{XX} \tag{A.1.47}$$

We now have all the pieces of equation (A.1.28) in equations (A.1.29), (A.1.33), and (A.1.47). Returning to equation (A.1.24), we see that we need $E(\mathbf{X}\widetilde{\mathbf{H}})$. Now equation (3.2.1.9) implies

$$E(\mathbf{X}\widetilde{\mathbf{H}}) = \mu_X - E\Big[S_{XX}\Sigma_{XX}^{-1}(\overline{\mathbf{X}}-\mu_{\mathrm{X}})\Big] = \mu_X \tag{A.1.48}$$

Now putting equation (A.1.28) together yields

$$\operatorname{cov}(\mathbf{X}\widetilde{\mathbf{H}}) = K^{-1}\Sigma_{XX} - 2K^{-1}\Sigma_{XX} + \frac{(K-1)(K+Q)}{(K-1)^2}\Sigma_{XX} \tag{A.1.49}$$

$$= \frac{Q+1}{(K-1)K}\Sigma_{XX} \; . \tag{A.1.50}$$

Now, insertion of equation (A.1.50) into equation (A.1.22) implies equation (A.1.18) becomes

$$\operatorname{cov}\Big[E\Big[\operatorname{vec}\mathbf{Y}\widetilde{\mathbf{H}} \,|\, \mathbf{Z} = z\Big]\Big] = \frac{Q+1}{(K-1)K}\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY} \; , \tag{A.1.51}$$

so equations (A.1.51) and (A.1.17) combine in equation (A.1.3) to yield equation (3.2.1.10).

Appendix 2: Derivation of Equation (3.2.1.15)

Let

$$c_1 = \left[\frac{K+Q-1}{K-1} \right]^p \quad \text{and} \quad c_2 = \left[\frac{K-2}{K+Q-1} \right] \tag{A.2.1}$$

We are interested in the conditions on ρ_i such that

$$\eta_2^* = c_1 \prod_{i=1}^{\nu} \left[1 - c_2 \rho_i^2 \right] < 1 \ . \tag{A.2.2}$$

Let $\nu = p$ and let ρ_* represent the smallest canonical correlation, now

$$\eta_2^* = c_1 \prod_{i=1}^{\nu} \left[1 - c_2 \rho_i^2 \right] < c_1 \prod_{i=1}^{\nu} \left[1 - c_2 \rho_*^2 \right] = \tag{A.2.3}$$

$$= c_1 \left[1 - c_2 \rho_*^2 \right]^{\nu} = \tag{A.2.4}$$

and

$$c_1 \left[1 - c_2 \rho_*^2 \right]^{\nu} < 1 \ \Rightarrow \ \rho_*^2 > \frac{1}{c_2} \left[1 - \left[\frac{1}{c_1} \right]^{1/\nu} \right], \tag{A.2.5}$$

substitution gives

$$\rho_*^2 > \left[\frac{K+Q-1}{K-2} \right] \left[1 - \left[\frac{K-1}{K+Q-1} \right]^{p/\nu} \right], \tag{A.2.6}$$

$$\rho_2^2 > \left(\frac{K+Q-1}{K-2} \right) \left[1 - \left(\frac{K-1}{K+Q-1} \right)^{p/\nu} \right], \tag{A.2.6}$$

but $\nu = p$ by assumption, so algebra reveals

$$\rho_2^2 > \left(\frac{K+Q-1}{K-2} \right) \left[1 - \left(\frac{K-1}{K+Q-1} \right) \right] = \frac{Q}{K-2} . \tag{A.2.7}$$

Appendix 3: Derivation of Equation (3.2.1.17)

Let

$$c(\nu,p) = \left[\frac{\dfrac{K+Q-1}{K-1}}{\dfrac{K-2}{K-Q-2}} \right]^{p/\nu} , \quad c_1 = \frac{K-2}{K+Q-1} . \tag{A.3.1}$$

Note $c(\nu,p)$, $c_1 < 1$, $\forall \nu$, $p > 0$. Write

$$\tilde{\eta} = \frac{\prod_{i=1}^{\nu} c(\nu,p)\left(1 - c_1\rho_i^2\right)}{\prod_{i=1}^{\nu}\left(1 - \rho_i^2\right)} \tag{A.3.2}$$

To start the induction let $\nu = 1$, then

$$\tilde{\eta} = \frac{c(1,p)\left(1 - c_1\rho_1^2\right)}{\left(1 - \rho_1^2\right)} < 1 , \tag{A.3.3}$$

implies

$$c(1,p)\left(1 - c_1\rho_1^2\right) < \left(1 - \rho_1^2\right) \Rightarrow \rho_1^2 > \frac{c(1,p)-1}{c(1,p)c_1 - 1} < 1 . \tag{A.3.4}$$

Hence conditions exist when $\nu = 1$. Now, when $\nu = n+1$

$$\tilde{\eta}(n+1) = \frac{\prod_{i=1}^{n} c(n+1,p)\left(1 - c_1\rho_i^2\right)}{\prod_{i=1}^{n}\left(1 - \rho_i^2\right)} \times \frac{c(n+1,p)\left(1 - c_1\rho_{n+1}^2\right)}{\left(1 - \rho_{n+1}^2\right)} \tag{A.3.5}$$

$$= \tilde{\eta}(n) \times \frac{c(n+1,p)\left[1 - c_1 \rho_{n+1}^2\right]}{\left[1 - \rho_{n+1}^2\right]} \tag{A.3.6}$$

Assume $\tilde{\eta}(n+1) < 1$, then

$$\tilde{\eta}(n) < \frac{c(n+1,p)\left[1 - c_1 \rho_{n+1}^2\right]}{\left[1 - \rho_{n+1}^2\right]} < 1 \,, \tag{A.3.7}$$

by an argument similar to that leading to equation (A.3.4). Now specify ρ_{n+1} as ρ_* and the induction follows.

Appendix 4: Derivation of Equation (4.2.4.1)

Now

$$E\left[X_j^R \mid N(t)\right] = 0 ,$$

and

$$\mathrm{var}\left[X_j^R \mid N(t)\right] = 1 ,$$

which implies

$$\mathrm{cov}\left[X_j^R, X_k^R \mid N(t)\right] = E\left[X_j^R X_k^R \mid N(t)\right] \tag{A.4.1}$$

for $j = 1, \cdots, S$ and for $k = 1, \cdots, S$. By substitution

$$\mathrm{cov}\left[X_j^R, X_k^R \mid N(t)\right] = \tag{A.4.2}$$

$$E\left[\sum_{l=1}^{N(t)} \frac{U_l(j) - p_j(*)}{\sqrt{N(t)(1-p_j(*))p_j(*)}} \sum_{m=1}^{N(t)} \frac{U_m(k) - p_k(*)}{\sqrt{N(t)(1-p_k(*))p_k(*)}} \;\middle|\; N(t)\right]$$

$$= \frac{1}{N(t)\sqrt{(1-p_j(*))p_j(*)(1-p_k(*))p_k(*)}} E\left[\sum_{l=1}^{N(t)} U_l(j) \sum_{m=1}^{N(t)} U_m(k)\right.$$

$$- p_j(*)N(t)\sum_{m=1}^{N(t)} U_m(k) - p_k(*)N(t)\sum_{l=1}^{N(t)} U_l(j) + p_j(*)p_k(*)N(t)^2 \mid N(t)\Bigg],$$

(A.4.3)

Now

$$E\left[p_j(*)N(t)\sum_{m=1}^{N(t)} U_m(k) \mid N(t)\right] = p_j(*)p_k(*)N(t)^2 \; . \qquad \text{(A.4.4)}$$

So insertion of equation (A.4.4) into the above yields

$$\operatorname{cov}\left[X_j^R, X_k^R \mid N(t)\right] =$$

$$\frac{E\left[\left(\sum_{l=1}^{N(t)} U_l(j) \sum_{m=1}^{N(t)} U_m(k)\right) - p_j(*)p_k(*)N(t)^2 \mid N(t)\right]}{N(t)\sqrt{(1-p_j(*))p_j(*)(1-p_k(*))p_k(*)}} \; .$$

(A.4.5)

Let $Y_j = \sum_{l=1}^{N(t)} U_l(j)$ and $Y_k = \sum_{m=1}^{N(t)} U_m(k)$. Now Y_j and Y_k are the marginals of a multinomial distribution. Hence

$$E\left[\sum_{l=1}^{N(t)} U_l(j) \sum_{m=1}^{N(t)} U_m(k) \mid N(t)\right] = E\left[Y_j Y_k \mid N(t)\right] \qquad \text{(A.4.6)}$$

$$= \operatorname{cov}\left[Y_j,Y_k \mid N(t)\right] + E\left[Y_j \mid N(t)\right]E\left[Y_k \mid N(t)\right] \qquad (A.4.7)$$

$$= -N(t)p_j(^*)p_k(^*) + N(t)^2 p_j(^*)p_k(^*)\ . \qquad (A.4.8)$$

So, equation (A.4.5) becomes (after some algebra),

$$\operatorname{cov}\left[X_j^R,X_k^R \mid N(t)\right] =$$

$$\frac{1}{N(t)\sqrt{(1-p_j(^*))p_j(^*)(1-p_k(^*))p_k(^*)}}\left[-N(t)p_j(^*)p_k(^*)\right] \quad (A.4.9)$$

$$= \frac{-p_j(^*)p_k(^*)}{\sqrt{(1-p_j(^*))p_j(^*)(1-p_k(^*))p_k(^*)}} \qquad (A.4.10)$$

$$= -\left[\frac{p_j(^*)p_k(^*)}{(1-p_j(^*))(1-p_k(^*))}\right]^{1/2} \qquad (A.4.11)$$

Appendix 5: Proof of Relation (4.2.4.2)

In this appendix we establish that

$$X^R \xrightarrow[n \to \infty]{D} N_s(\mathbf{0}, \Sigma_{RR}) \tag{A.5.1}$$

where $X^R = \left[X_1^R, \cdots, X_s^R\right]$. Here there are $1, \cdots, s$ stations with positive probability of being branched to from the CPU. We closely follow a similar proof offered by Wilson and Pritsker (1984a, appendix A).

Let $g(t)$ be the greatest integer in $\alpha_k t$ for $t \geq 0$. Here

$$\alpha_k = \lim_{t \to \infty} \rho_{CPU} \frac{p_k(*)}{E\left[U_{CPU}\right]} \quad \text{with probab}\text{'' }\text{ty } 1 \tag{A.5.2}$$

where k is the station branched to, ρ_{CPU} is the utilization of the CPU, and U_{CPU} is the service time distribution of the CPU. Now express the standardized routing variable $X_k^R(t)$ in terms of the partial sum process

$$S_n(k) = \sum_{j=1}^{n} \frac{U_j(k) - \mu_k}{\sigma_k}, \quad n \geq 1. \tag{A.5.3}$$

Wilson and Pritsker's proof is based on use of the "dissection" formula

$$X_k^R(t) = Z_k(t) + F_k(t) + R_k(t) \tag{A.5.4}$$

where

$$Z_k(t) = S_{g(k,t)}/g(k,t)^{1/2}$$

$$F_k(t) = \left\{ \left[g(k,t)/N(k,t) \right]^{1/2} - 1 \right\} Z_k(t)$$

$$R_k(t) = \left[g(k,t)/N(k,t) \right]^{1/2} \left\{ \left[S_{N(k,t)} - S_{g(k,t)} \right] / g(k,t)^{1/2} \right\}.$$

(A.5.6)

In vector notation we set $Z(t) = \left[Z_1(t), \cdots, Z_s(t) \right]'$, $F(t) = \left[F_1(t), \cdots, F_s(t) \right]'$, $R(t) = \left[R_1(t), \cdots, R_s(t) \right]'$.

$$X^R(t) = Z(t) + F(t) + R(t) \qquad \text{(A.5.7)}$$

To prove equation (A.5.1) we show that, given $Z \sim N_s(\mathbf{0}, \Sigma_{RR}^{(s-1)})$

$$b'X^R(t) \xrightarrow[t \to \infty]{D} b'Z \quad \forall b \in E^s . \qquad \text{(A.5.8)}$$

Here, $\Sigma_{RR}^{(s-1)}$ is the covariance matrix of standardized controls with one of the controls removed (more on this below). To show this we examine the

asymptotic behavior of each component of $b'X^R(t)$. First, we consider the component $b'Z(t)$. We compute the moment generating function of $b'Z(t)$:

$$\mathrm{M}_{b'Z(t)}(\theta) = E\left[\exp\left\{\theta b'Z(t)\right\}\right]$$

$$= E\left[\exp\left\{(\theta b)'Z(t)\right\}\right]$$

$$= \mathrm{M}_{Z(t)}(b\,\theta) \qquad \text{(A.5.9)}$$

Now the multivariate central limit theorem (Neuts (1973), pg 287) insures

$$Z(t) \xrightarrow[n \to \infty]{D} N_s(\mathbf{0}, \Sigma_{RR}^{(s-1)}) \;,\; |\Sigma_{RR}^{(s-1)}| > 0 \qquad \text{(A.5.10)}$$

In the next two paragraphs we demonstrate that $|\Sigma_{RR}^{(s-1)}| > 0$.

First, notice that if X is a $p \times 1$ vector of random variables with positive definite covariance matrix Σ_{XX}, then if we transform X as

$$Z = \mathrm{diag}(\sigma_1^{-1}, \cdots, \sigma_k^{-1})\, X \qquad \text{(A.5.11)}$$

$$\mathrm{cov}[Z] = \mathrm{diag}(\sigma_1^{-1}, \cdots, \sigma_k^{-1})\, \mathrm{cov}[X]\, \mathrm{diag}(\sigma_1^{-1}, \cdots, \sigma_k^{-1}) \qquad \text{(A.5.12)}$$

$$\text{cov}[Z] = \left\{ \prod_{i=1}^{p} \sigma_i^{-2} \right\} \text{cov}[X] > 0 , \tag{A.5.13}$$

hence the covariance matrix of the standardized random varibles is also positive definite.

For notational convenience let $N(t) = N$, $\sum_{l=1}^{N} U_l(\mathrm{k}) = N_{\mathrm{k}}$, and $\mathrm{N} = \left[N_1, \cdots, N_s \right]$ then

$$\sum_{k=1}^{s} N_{\mathrm{k}} = N \rightarrow |\Sigma_{NN}^{(s)}| = 0. \tag{A.5.14}$$

Now

$$|\Sigma_{NN}^{(s-1)}| = 0 \rightarrow \exists\ \lambda_1, \cdots, \lambda_s \text{ with some } \lambda_i \neq 0 \tag{A.5.15}$$

such that

$$\sum_{k=1}^{s-1} \lambda_{\mathrm{k}} N_{\mathrm{k}} = 0 \quad \text{w.p. } 1 \tag{A.5.16}$$

If we add (A.5.14) and (A.5.16) we get

$$\sum_{k=1}^{s-1} (1+\lambda_{\mathrm{k}}) N_{\mathrm{k}} + N_s = N \quad \text{w.p. } 1 \tag{A.5.17}$$

Now for all k, $1 \leq \mathrm{k} \leq s-1$

$$p_k(*) > 0 \longrightarrow \Pr\left\{N_k, N_j{=}0, j \neq s\right\} > 0 \qquad \text{(A.5.18)}$$

$$\longrightarrow (1+\lambda_k)N = N \qquad \text{(A.5.19)}$$

$$\longrightarrow \lambda_k = 0\ , \qquad \text{(A.5.20)}$$

and since $p_k(*) > 0\ ,\ \forall\ k$ we have a contradiction. Therefore, the covariance matrix of the non-standardized (hence the standardized) controls is positive definite. Now $|\Sigma_{RR}^{(s-1)}| > 0$ validates (A.5.10) and this implies

$$\lim_{t \longrightarrow \infty} \mathrm{M}_{Z(t)}(\theta b) = \mathrm{M}_Z(\theta b)$$

$$= E\left[\exp\left\{(\theta b)'Z\right\}\right]$$

$$= E\left[\exp\left\{\theta b'Z\right\}\right]$$

$$= \mathrm{M}_{b'Z}(\theta) \qquad \text{(A.5.21)}$$

and equation (A.5.8) follows. Using nearly identical arguments to Wilson and Pritsker, we can show that

$$b'F(t) \xrightarrow[t \to \infty]{D} \mathbf{0} \, , \; b'R(t) \xrightarrow[t \to \infty]{D} \mathbf{0} \, . \tag{A.5.22}$$

To finish the proof we apply Slutsky's theorem twice and invoke the Cramer-Wold theorem (implied in equation (A.5.8)).

Appendix 6: FORTRAN Listings of SLAM Models

```
c      program main(input,output,tape5=input,tape6=output,tape7,tape1,
c     1tape2,tape3,tape4)
      program main
      dimension nset(5000)
      common qset(5000)
      common/scom1/ atrib(100),dd(100),ddl(100),dtnow,ii,mfa,mstop,nclnr
     1,ncrdr,nprnt,nnrun,nnset,ntape,ss(100),ssl(100),tnext,tnow,xx(100)
      common/ucom1/ depart(5),rmean(4),p(4,4),servt(4),ecount(2)
      equivalence (nset(1),qset(1))
      nnset=5000
      ncrdr=5
      nprnt=6
      ntape=7

      read (ncrdr,*) (rmean(i),i=1,4)

      do 12 i=1,4
       read (ncrdr,*) (p(i,j),j=1,4)
12     continue

      call slam
      stop
      end
c
c***********************************************************************
c
      subroutine event(i)
      common/scom1/ atrib(100),dd(100),ddl(100),dtnow,ii,mfa,mstop,nclnr
     1,ncrdr,nprnt,nnrun,nnset,ntape,ss(100),ssl(100),tnext,tnow,xx(100)
      common/ucom1/ depart(5),rmean(4),p(4,4),servt(4),ecount(2)

      ecount(1)=ecount(1)+1
      if(tnow.gt.2000) ecount(2)=ecount(2)+1

      goto (1,2),i

1      call arss
      return
2      call endss
      return
      end
c
c***********************************************************************
c
      subroutine intlc
      common/scom1/ atrib(100),dd(100),ddl(100),dtnow,ii,mfa,mstop,nclnr
     1,ncrdr,nprnt,nnrun,nnset,ntape,ss(100),ssl(100),tnext,tnow,xx(100)
      common/ucom1/ depart(5),rmean(4),p(4,4),servt(4),ecount(2)
      common/gcom5/ iised(10),jjbeg,jjclr,mmnit,mmon,nname(5),nncfi,
     &nnday,nnpt,nnprj(5),nnrns,nnstr,nnyr,sseed(10),lseed(10)
      integer iseed(1000)
      common/ucom2/ multino(4)
```

```
      if(nnrun.eq.1) then
       do 1 i=1,1000
        iseed(i)=(1.e+12)*drand(1)
1       continue
      endif
      iised(2)=iseed(nnrun)
      x=drand(-2)

      do  4 i=1,4
       multino(i)=0
4     continue

      do  5 i=1,2
       ecount(i)=0.
5     continue

      do 6 i=1,5
       depart(i)=0.
6     continue

      do 7 i=1,4
       servt(i)=0.
7     continue

      do 10 i=1,25
       etime=expon(rmean(1),2)
       atrib(1)=etime
       atrib(3)=i
       atrib(4)=1
       atrib(5)=2
       call schdl(i,etime,atrib)
10     continue

      do 11 i=1,4
       xx(i)=0.
11     continue
      write(6,99)nnrun
99     format(1x,'SIMULATION STUDY IN PROGRESS : RUN ',i4, ' OF
     &1000 RUNS')
      return
      end
c
c*************************************************************************
c
      subroutine endss
      common/scom1/ atrib(100),dd(100),ddl(100),dtnow,ii,mfa,mstop,nclnr
     1,ncrdr,nprnt,nnrun,nnset,ntape,ss(100),ssl(100),tnext,tnow,xx(100)
      common/ucom1/ depart(5),rmean(4),p(4,4),servt(4),ecount(2)
      common/ucom2/ multino(4)

      call schdl(1,0.,atrib)
```

```
      myq=atrib(4)

      if(nnq(myq).ne.0) then
        call rmove(1,myq,atrib)
        wait=tnow-atrib(2)
        call colct(wait,myq)
        rm=rmean(myq)
        service=expon(rm,2)
        atrib(4)=atrib(5)
        iat=atrib(4)+.00001
        call nextguy(iat,inext)
c
c COLLECT STATISTICS WHILE PARKED AT CPU
c
        if(iat.eq.2) then
         multino(inext)=multino(inext)-1
        endif

        atrib(5)=inext
        call schdl(2,service,atrib)
        if(tnow.gt.2000) then
          servt(myq)=servt(myq)+service
          depart(myq)=depart(myq)+1
          depart(5)=depart(5)+1
        endif
      else
        xx(myq)=0.
      endif

      return
      end
c
c***********************************************************************
c
      subroutine arss
      common/scom1/ atrib(100),dd(100),ddl(100),dtnow,ii,mfa,mstop,nclnr
     1,ncrdr,nprnt,nnrun,nnset,ntape,ss(100),ssl(100),tnext,tnow,xx(100)
      common/ucom1/ depart(5),rmean(4),p(4,4),servt(4),ecount(2)
      common/ucom2/ multino(4)

      iat=atrib(5)

      if(iat.eq.1) then
        resp=tnow-atrib(1)
        call colct(resp,1)


        rm=rmean(1)
        service=expon(rm,2)
        atrib(1)=tnow+service
        atrib(4)=1
        atrib(5)=2
```

```
        call schdl(1,service,atrib)
        if(tnow.gt.2000) servt(iat)=servt(iat)+service
      else
        if(xx(iat).gt.0.) then
          atrib(2)=tnow
          call filem(iat,atrib)
          return
        else
          wait=0.
          call colct(wait,iat)
          rm=rmean(iat)
          atrib(4)=iat
          call nextguy(iat,inext)
c
c COLLECT STATISTICS WHILE PARKED AT CPU
c
        if(iat.eq.2) then
         multino(inext)=multino(inext)+1
        endif

          atrib(5)=inext
          service=expon(rm,2)
          xx(iat)=1
          call schdl(2,service,atrib)
          if(tnow.gt.2000) servt(iat)=servt(iat)+service
        endif
      endif

      if (tnow.gt.2000) then
        depart(iat)=depart(iat)+1
        depart(5)=depart(5)+1
      endif

      return
      end
c
c*******************************************************************************
c
      subroutine nextguy(iat,inext)
      common/ucom1/ depart(5),rmean(4),p(4,4),servt(4),ecount(2)

      cum=0.
      u=unfrm(0.,1.,2)

      do 10 index=1,4
       cum=cum+p(iat,index)
       if(u.le.cum) then
         inext=index
         goto 11
       else
         continue
       endif
```

```
10    continue

11    return
      end

************************************************************************

      subroutine otput
      common/scom1/ atrib(100),dd(100),ddl(100),dtnow,ii,mfa,mstop,nclnr
     1,ncrdr,nprnt,nnrun,nnset,ntape,ss(100),ssl(100),tnext,tnow,xx(100)
      common/ucom1/ depart(5),rmean(4),p(4,4),servt(4),ecount(2)
      common/ucom2/ multino(4)

      write(1,*)nnrun
      write(1,*)(ecount(i),i=1,2)
      write(1,*)(ccavg(i),i=1,4)
      write(1,*)(ttavg(i),i=2,4)
      write(1,*)(servt(i),i=1,4)
      write(1,*)(depart(i),i=1,5)
      write(1,*)(ffawt(i),i=2,4)
      isum=0
      do 1 i=1,4
       isum=isum+multino(i)
1     continue
      write(1,*)(multino(i),i= 1,4),isum

      return
      end
```

```
c      program main(input,output,tape5=input,tape6=output,tape7,tape1,
c     1tape2,tape3,tape4)
      program main
      dimension nset(5000)
      common qset(5000)
      common/scom1/ atrib(100),dd(100),ddl(100),dtnow,ii,mfa,mstop,nclnr
     1,ncrdr,nprnt,nnrun,nnset,ntape,ss(100),ssl(100),tnext,tnow,xx(100)
      common/ucom1/ depart(10),rmean(10),p(10,10),servt(10),ecount(2)
      common/ucom2/ isubcap,nusssn,numcust,tclear,nstudy
      equivalence (nset(1),qset(1))
      nnset=5000
      ncrdr=5
      nprnt=6
      ntape=7

      read (ncrdr,*) isubcap,nusssn,numcust,tclear,nstudy
      read (ncrdr,*) (rmean(i),i=1,nusssn-2)

      do 12 i=1,nusssn-2
       read (ncrdr,*) (p(i,j),j=1,nusssn-2)
12     continue

      call slam
      stop
      end
c
c*********************************************************************
c
      subroutine event(i)
      common/scom1/ atrib(100),dd(100),ddl(100),dtnow,ii,mfa,mstop,nclnr
     1,ncrdr,nprnt,nnrun,nnset,ntape,ss(100),ssl(100),tnext,tnow,xx(100)
      common/ucom1/ depart(10),rmean(10),p(10,10),servt(10),ecount(2)
      common/ucom2/ isubcap,nusssn,numcust,tclear,nstudy

      ecount(1)=ecount(1)-1
      if(tnow.gt.tclear) ecount(2)=ecount(2)-1

      goto (1,2),i

1      call arss
      return
2      call endss
      return
      end
c
c*********************************************************************
c
      subroutine intlc
      common/scom1/ atrib(100),dd(100),ddl(100),dtnow,ii,mfa,mstop,nclnr
     1,ncrdr,nprnt,nnrun,nnset,ntape,ss(100),ssl(100),tnext,tnow,xx(100)
      common/ucom1/ depart(10),rmean(10),p(10,10),servt(10),ecount(2)
      common/ucom2/ isubcap,nusssn,numcust,tclear,nstudy
```

```
      common/gcom5/ iised(10),jjbeg,jjclr,mmnit,mmon,nname(5),nncfi,
     &nnday,nnpt,nnprj(5),nnrns,nnstr,nnyr,sseed(10),lseed(10)
      common/ucom3/ multino(7)
      integer iseed(2000)

      if(nnrun.eq.1) then
       do 1 i=1,2000
         iseed(i)=(1.e+12)*drand(1)
1       continue
      endif
      iised(2)=iseed(nnrun)
      x=drand(-2)

      do  4 i=1,7
       multino(i)=0
4     continue

      do  5 i=1,2
       ecount(i)=0.
5     continue

      do 6 i=1,nusssn+3
       depart(i)=0.
6     continue

      do 7 i=1,nusssn+2
       servt(i)=0.
7     continue

      do 10 i=1,numcust
       etime=expon(rmean(1),2)
       atrib(1)=etime
       atrib(3)=i
       atrib(4)=1
       atrib(5)=2
       call schdl(1,etime,atrib)
10     continue

      do 11 i=1,nusssn+2
       xx(i)=0.
11     continue
      write(6,99)nnrun,nstudy
99     format(1x,'SIMULATION STUDY IN PROGRESS : RUN ',i4, ' OF
     &'i4,' RUNS')
      return
      end
c
c*******************************************************************************
c
      subroutine endss
      common/scom1/ atrib(100),dd(100),ddl(100),dtnow,ii,mfa,mstop,nclnr
     1,ncrdr,nprnt,nnrun,nnset,ntape,ss(100),ssl(100),tnext,tnow,xx(100)
```

```
      common/ucom1/ depart(10),rmean(10),p(10,10),servt(10),ecount(2)
      common/ucom2/ isubcap,nusssn,numcust,tclear,nstudy
      common/ucom3/ multino(7)

      call schdl(1,0.,atrib)
      myq=atrib(4)

      if(nnq(myq).ne.0) then
        call rmove(1,myq,atrib)
        wait=tnow-atrib(2)
        call colct(wait,myq)
        rm=rmean(myq)
        service=expon(rm,2)
        atrib(4)=atrib(5)
        iat=atrib(4)+.00001
        call nextguy(iat,inext)
c
c COLLECT STATISTICS WHILE PARKED AT CPU
c
        if(iat.eq.3) then
         multino(inext)=multino(inext)+1
        endif

        atrib(5)=inext
        call schdl(2,service,atrib)
        if(tnow.gt.tclear) then
          servt(myq)=servt(myq)+service
          depart(myq)=depart(myq)+1
          depart(nusssn+3)=depart(nusssn+3)+1
        endif
      else
        xx(myq)=0.
      endif

      if(myq.eq.3.and.nnq(2).gt.0.and.isubcap.ne.0.and.inext.eq.1
     &.and.nnq(myq).ne.0) then
        call rmove(1,2,atrib)
        service=0.
        atrib(4)=atrib(5)
        atrib(5)=3
        call schdl(1,service,atrib)
      endif

      return
      end
c
c***********************************************************************
c
      subroutine arss
      common/scom1/ atrib(100),dd(100),ddl(100),dtnow,ii,mfa,mstop,nclnr
     1,ncrdr,nprnt,nnrun,nnset,ntape,ss(100),ssl(100),tnext,tnow,xx(100)
      common/ucom1/ depart(10),rmean(10),p(10,10),servt(10),ecount(2)
```

```
      common/ucom2/ isubcap,nusssn,numcust,tclear,nstudy
      common/ucom3/ multino(7)

      iat=atrib(5)

      if(iat.eq.1) then
        resp=tnow-atrib(1)
        call colct(resp,1)
        rm=rmean(1)
        service=expon(rm,2)
        atrib(1)=tnow+service
        atrib(4)=1
        atrib(5)=2
        call schdl(1,service,atrib)
        if(tnow.gt.tclear) servt(iat)=servt(iat)+service
        go to 101
      endif

      if(iat.eq.2) then
        if(isubcap.ne.0) then
          numsub=0
          do 10 i=3,nusssn+2
           numsub=numsub+nnq(i)+xx(i)
10         continue
          if(numsub.lt.isubcap) then
            if(nnq(2).eq.0) then
              wait=0.
              call colct(wait,2)
              service=0.
              atrib(4)=2
              atrib(5)=3
              call schdl(1,service,atrib)
              go to 101
            else
              atrib(2)=tnow
              call filem(2,atrib)
              call rmove(1,2,atrib)
              wait=tnow-atrib(2)
              call colct(wait,2)
              atrib(4)=2
              atrib(5)=3
              service=0.
              call schdl(1,service,atrib)
              go to 101
            endif
          else
            atrib(2)=tnow
            call filem(2,atrib)
            return
          endif
        endif
      endif
```

```
100    if(xx(iat).gt.0.) then
        atrib(2)=tnow
        call filem(iat,atrib)
        return
      else
        wait=0.
        call colct(wait,iat)
        rm=rmean(iat)
        atrib(4)=iat
        call nextguy(iat,inext)
c
c COLLECT STATISTICS WHILE PARKED AT CPU
c
      if(iat.eq.3) then
       multino(inext)=multino(inext)+1
      endif

        atrib(5)=inext
        service=expon(rm,2)
        xx(iat)=1
        call schdl(2,service,atrib)
        if(tnow.gt.tclear) servt(iat)=servt(iat)+service
      endif

101   if (tnow.gt.tclear) then
       depart(iat)=depart(iat)+1
       depart(nusssn+3)=depart(nusssn+3)+1
     endif

     return
     end
c
c***********************************************************************
c
     subroutine nextguy(iat,inext)
     common/ucom1/ depart(10),rmean(10),p(10,10),servt(10),ecount(2)
     common/ucom2/ isubcap,nusssn,numcust,tclear,nstudy

     cum=0.
     u=unfrm(0.,1.,2)

     do 10 index=1,nusssn+2
      cum=cum+p(iat,index)
      if(u.le.cum) then
        inext=index
        goto 11
      else
        continue
      endif
10    continue

11    return
```

```
      end
*****************************************************************

*****************************************************************

      subroutine otput
      common/scom1/ atrib(100),dd(100),ddl(100),dtnow,ii,mfa,mstop,nclnr
     1,ncrdr,nprnt,nnrun,nnset,ntape,ss(100),ssl(100),tnext,tnow,xx(100)
      common/ucom1/ depart(10),rmean(10),p(10,10),servt(10),ecount(2)
      common/ucom2/ isubcap,nusssn,numcust,tclear,nstudy
      common/ucom3/ multino(7)

      write(1,*)nnrun
      write(1,*)(ecount(i),i=1,2)
      write(1,*)(ccavg(i),i=1,nusssn+2)
      write(1,*)(ttavg(i),i=2,nusssn+2)
      write(1,*)(servt(i),i=1,nusssn+2)
      write(1,*)(depart(i),i=1,nusssn+3)
      isum=0
      do 1 i=1,7
       isum=isum+multino(i)
1      continue
      write(1,*)(multino(i),i=1,7),isum


      return
      end
```

Appendix 7: FORTRAN Listing of the Analysis Program

```
c    program tree(input,output,tape7,tape5=input,tape6=output)
*******************************************************************
*                                                                 *
*                                                                 *
*   This program uses an "all possible regressions" approach to   *
*   select the best subset of controls from a given candidate set. *
*   It assumes that certain number of meta-experiments have been  *
*   performed each with the same number of replications. Once the *
*   optimal subset has been identified, a confidence region is    *
*   constructed about the mean vector for the responses. Coverage *
*   and volume reduction is tallyed and subsequently summarized.  *
*                                                                 *
*   The program can be run in two modes. The user can either      *
*   estimate the covariance matrix of controls or incorporate     *
*   it directly. The program variable "iknow" dictates which      *
*   option is in effect (see code below).                         *
*                                                                 *
*   The program can also be run in the "best m" regressions mode. *
*   ( Currently only configured for estimated covariance matrix   *
*     of controls)                                                *
*   In other words it will compute the best m subsets of each     *
*   possible subset size. This can be of interest if a single set *
*   of data is used.                                              *
*                                                                 *
*                                                                 *
*   PARAMETERS TO BE INITIALIZED:                                 *
*                                                                 *
*      nx      = # of candidate controls                          *
*      ny      = # of responses                                   *
*      keepers = # of best regressions to be kept                 *
*                (m in "m best" as above)                         *
*      numreps = # replications per meta experiment               *
*      meta    = # of meta experiments                            *
*                                                                 *
*  NOTE:                                                          *
*                                                                 *
*      IN SUBROUTINE COVER : nx2 AND ny2 MUST BE SET              *
*                    TO nx and ny RESPECTIVELY.                   *
*                    (IN THE PARAMETER STATEMENT)                 *
*                                                                 *
*******************************************************************
      program tree
      parameter (nx=7,ny=2,nvar=nx+ny,keepers=6,knx=2**nx)
      parameter (numreps=20,meta=50)
      parameter(n1=nvar,n2=numreps,n3=numreps,n4=1,n5=1,n6=0)
      parameter(nn1=((n1*(n1+1))/2))
      common sig,kk,iqq,ip
      character*25 title,respons(ny),control(nx)
      real a(nvar,nvar,nvar)
      real wkarea(ny),rss(ny,ny),dum(ny)
      integer nk(nvar)
      dimension x(n3,n1),nbr(6),temp(n1),xm(n1)
```

```
      real vcv((n1*(n1-1))/2),full(n1,n1)
      real regr(keepers,nx,2),buff(keepers),buff2(keepers)
      real ff(0:nx)
      external f
      integer models(knx,nx),ibuff(nx)
      integer ih(n1)
      integer icover(4),ictot(4)
      real vecybar(ny),vr(2),volred(2),coverag(4)
      real vecmuy(ny),vecmuc(nx),ybar(ny),cbar(nx),veccbar(nx)
      real covcv(nx,nx)
      real target(ny,ny)

      data vecmuc/0.,0.,0.,0.,0.,0.,0./
      data vecybar /78.31305,.4132402/
      data vecmuy /81.71,.413/
      data veccbar/-2.169668E-02,-1.416941E-02,5.544987E-02,
     &           -1.809913E-02
     &           ,3.908565E-02,-1.957430E-02,3.610350E-03/
      data title/'MODEL5:TRANSFORMED'/
      data respons/'SYSTEM RESPONSE TIME',
     &          'CPU UTILIZATION     '/
      data control/'ROUTING VARIABLE (1)',
     &          'ROUTING VARIABLE (3)',
     &          'ROUTING VARIABLE (4)',
     &          'WORK VARIABLE (1)',
     &          'WORK VARIABLE (2)',
     &          'WORK VARIABLE (3)',
     &          'WORK VARIABLE (4)'/

      open (unit=1,file='out.model.5',status='new')
      write(1,31) title,meta,numreps,meta*numreps
31     format(1x,a25,'meta = ',i3,' numreps = ',i3,' total reps = ',
     &i4)
      write(1,32)meta*numreps
32     format(1x,'the response are',13x,'mean ',i4,' reps',2x,
     &'steady state mean'/)
      do 33 i=1,ny
       write(1,34)i,respons(i),vecybar(i),vecmuy(i)
34      format(2x,i2,1x,a25,f12.5,4x,f12.5)
33     continue
      write(1,35)
35     format(' ')
      write(1,36)meta*numreps
36     format(1x,'the candidate controls are',3x,'mean ',i4,' reps',
     &2x,'steady state mean'/)
      do 37 i=1,nx
       write(1,34)i,control(i),veccbar(i),vecmuc(i)
37     continue
      write(1,35)

c     iknow IS THE FLAG FOR USE OF THE KNOWN COVARIANCE MATRIX
c                         OF CONTROLS
```

```
c
c              iknow = 1 known cov used
c              iknow = 0 cov estimated
c
      iknow=0

      if(iknow.eq.0) then
       write(1,38)
      else
       write(1,39)
      endif
38     format(/,1x,'COVARIANCE MATRIX OF CONTROLS WAS ESTIMATED')
39     format(/,1x,'KNOWN COVARIANCE MATRIX OF CONTROLS WAS USED')
c
c     HERE WE READ THE KNOWN COVARIANCE STRUCTURE OF CONTROLS
c              (IF REQUIRED)
c
      if(iknow.eq.1) then
       open(unit=3,file='cov.model.5',status='old')
       rewind 3
       do 21 i=1,nx
        read(3,*)(covcv(i,j),j=1,nx)
21      continue
      endif

      nbr(1)=n1
      nbr(2)=n2
      nbr(3)=n3
      nbr(4)=n4
      nbr(5)=n5
      nbr(6)=n6
      ix=n3
      sig=.90
c
c    MAKE THE F TABLE
c
       ip=ny
       kk=numreps
       call ftabl(ff,nx)
       print *,'f table',ff

c
c       iwrite = 0  Meta Experiment mode
c       iwrite = 1  Best m Regressions mode
c                (m=keepers above)(meta = 1)

      iwrite=0

c
c     INITIALIZE COVERAGE AND VOLUME REDUCTION ACCUMULATORS
c
```

```
      do 850 iz=1,4
       ictot(iz)=0
850    continue
      do 851 iz=1,2
       vr(iz)=0.
851    continue
c
c   THIS IS THE META EXPERIMENT LOOP
c
      do 1000 mm=1,meta

c
c     INITIALIZE ARRAYS
c

      numreg=0
      do  999 iz=1,keepers
      do  999 jz=1,nx
      do  999 kz=1,2
       regr(iz,jz,kz)=0.
999    continue
      do 998 iz=1,knx
      do 998 jz=1,nx
       models(iz,jz)=0
998    continue
      do 997 iz=1,nvar
      do 997 jz=1,nvar
      do 997 kz=1,nvar
       a(iz,jz,kz)=0.
997    continue
      do 996 iz=1,keepers
       buff(iz)=0
       buff2(iz)=0
996    continue
      do 995 iz=1,nx
       ibuff(iz)=0
995    continue

c
c     READ THE DATA (each record => [controls|responses])
c     COMPUTE THE COVARIANCE MATRIX
c     SAVE SAMPLE MEANS
c     BOUND THE GENERALIZED VARIANCE
c

      do 10 i=1,n2
       read(5,*)(x(i,j),j=1,n1)
10     continue
      call becovm(x,ix,nbr,temp,xm,vcv,ier)
      do 13 i=1,nx
       cbar(i)=xm(i)
13     continue
```

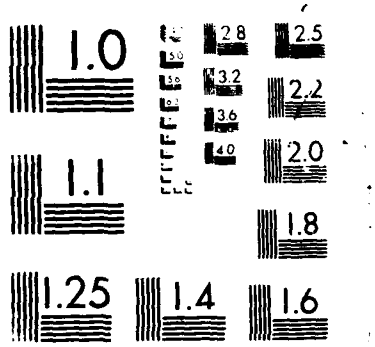
```
      do 14 i=nx+1,n1
       ybar(i-nx)=xm(i)
14     continue
      call vcvtsf(vcv,n1,full,n1)
      do 11 i=1,n1
      do 11 j=1,n1
       a(1,i,j)=full(i,j)
11     continue
77     is=1
      do 99 ii=1,ny
      do 99 jj=1,ny
       if(jj.gt.ii) then
        rss(ii,jj)=a(is,nx+ii,nx+jj)
        rss(jj,ii)=rss(ii,jj)
       else
        if(ii.eq.jj) then
         rss(ii,jj)=a(is,nx+ii,nx+jj)
        endif
       endif
99     continue
      iopt=5
      call linv3f(rss,dum,4,ny,ny,d1,d2,wkarea,ier)
      if(ier.ne.0)print *,"IDIED BELOW 99"
      det=d1*2**d2
      big=(float(numreps-1)/float(numreps-nx-2))**ny
      two=2*big*d1*2**d2
c
c   STUFF THE BOOKKEEPING ARRAY WITH THE BOUND
c
      do 200 ii=1,keepers
      do 200 jj=1,nx
       regr(ii,jj,1)=two
200    continue
c
c   CONDUCT A BINARY SEARCH OF THE REGRESSION TREE
c          FURNIVAL AND WILSON (1974)
c

      k=nx
      do 1 l=1,k
       nk(l)=0
1     continue
      nk(k+1)=1
      l=1
2     nk(l)=1
      do 3 m=l,k
       if(nk(m+1).eq.1) go to 4
3     continue
4     call gauss(k-m+1,k-l+2,k-l+1,a,nvar,nvar)

c
c     CALCULATION OF THE GENERALIZED RESIDUAL COVARIANCE
```

```
c
      is=k-l+2
      do 100 ii=1,ny
      do 100 jj=1,ny
       if(jj.gt.ii) then
        rss(ii,jj)=a(is,nx+ii,nx+jj)
        rss(jj,ii)=rss(ii,jj)
       else
        if(ii.eq.jj) then
         rss(ii,jj)=a(is,nx+ii,nx+jj)
        endif
       endif
100    continue
      iopt=5


      if(iknow.eq.0) then
       call linv3f(rss,dum,4,ny,ny,d1,d2,wkarea,ier)
      if(ier.ne.0)print *,"IDIED BELOW 100"
       det=d1*2**d2
      endif

c
c   BOOKKEEPING LOGIC TO SAVE M=KEEPERS BEST REGRESSIONS
c   OF ALL J SUBSETS SIZES
c
      mv=o
      do 300 n=1,nx
       mv=mv+nk(n)
300    continue

      if(iknow.eq.0) then
       const=(float(numreps-1)/float(numreps-mv-1))
       det=det*const**ny
      else
       call covknow(rss,ny,full,nvar,target,dum,numreps,mv,det)
      endif

      do 301 j=1,keepers
       if(det.lt.regr(j,mv,1)) then
         numreg=numreg+1
        do 302 jj=j,keepers-1
          buff(jj+1)=regr(jj,mv,1)
          buff2(jj+1)=regr(jj,mv,2)
302       continue
         regr(j,mv,1)=det
         regr(j,mv,2)=numreg
        do 303 jj=j+1,keepers
          regr(jj,mv,1)=buff(jj)
          regr(jj,mv,2)=buff2(jj)
303       continue
         call keepit(numreg,nk,nx,models,knx,nvar)
```

```
      go to 304
     endif
301   continue
304   continue
     do 5 l=1,k
      if(nk(l).eq.0) go to 2
      nk(l)=0
5    continue
c
c    THIS BLOCK IS FOR BEST M SUBSETS MODE OF OPERATION
c
     if(iwrite.eq.1) then
     do 500 i=1,nx
      write(1,600) keepers,i
600    format(10x,'best ',i2,' regressions with ',i2,' variables'//)
     do 500 j=1,keepers
      ivar=0
      iin=0
      do 400 ii=nx,1,-1
       ivar=ivar+1
       if(ifix(regr(j,i,2)+.0001).eq.0) go to 500
       if(models(ifix(regr(j,i,2)+.0001),ii).eq.1) then
        iin=iin+1
        ibuff(iin)=ivar
       endif
400    continue
      rdet=regr(j,i,1)
      write(1,601)rdet,(ibuff(ij),ij=1,iin)
601    format(1x,e16.8,10x,30(i2,1x))
500   continue
     endif

c
c    FOR EACH SUBSET COMPUTE THE CRITERION AND SAVE THE MINIMUM
c
     if(iwrite.eq.0) then
      ip=ny
      kk=numreps
      do 650 iq=1,nx

       if(iknow.eq.0) then
        regr(1,iq,1)=regr(1,iq,1)*c3(kk,iq,ip)*cfront(kk,iq,ip)*
    &    ff(iq)
       else
        regr(1,iq,1)=regr(1,iq,1)*c4(kk,iq,ip)*cfront(kk,iq,ip)*
    &    ff(iq)
       endif

       if(iq.eq.1) rmin=regr(1,iq,1)+1000.
650    continue
      do 700 iq=1,nx
       if(regr(1,iq,1).lt.rmin) then
```

```
        rmin=regr(1,iq,1)
        iat= regr(1,iq,2)
       endif
700    continue
      ivar=0
      iin=0
      do 750 ii=nx,1,-1
       ivar=ivar+1
       if(models(iat,ii).eq.1) then
        iin=iin+1
        ibuff(iin)=ivar
       endif
750    continue
      sp=rmin
      write(1,601)sp,(ibuff(ij),ij=1,iin)
c
c   FIND THE VOLUME REDUCTION AND INDICATE COVERAGE
c
      call cover(vcv,n1,nn1,models,knx,nx,iat,iin,ybar
     &,cbar,vecmuc,ny,vecmuy,numreps,ff,ih,icover,volred,vecybar
     &,iknow,covcv)
c
c    COVERAGE AND VOLUME REDUCTION TALLYS
c
      do 800 ic=1,4
       ictot(ic)=ictot(ic)+icover(ic)
800   continue
      do 801 ic=1,2
       vr(ic)=vr(ic)+volred(ic)
801   continue

      endif
      print *,"THIS IS META-EXPERIMENT # ",mm,"icover ",icover

1000  continue

      do 1001 iz=1,2
       vr(iz)=vr(iz)/float(meta)
1001  continue
      do 1002 iz=1,4
       coverag(iz)=float(ictot(iz))/float(meta)
1002  continue

      write(1,602)coverag(1),vr(1)
602   format(1x,'contrld coverage on steady [illegible]
     &' vol reduct ',e16.8)
      write(1,603)coverag(2)
603   format(1x,'uncontrld coverage [illegible]
      write(1,604)coverag(3),vr(1)
604   format(1x,'contrld coverage [illegible]
     &' vol reduct ',e16.8)
      write(1,605)coverag(4
```

AD-A186 637 CONTROL VARIATE SELECTION FOR MULTIRESPONSE SIMULATION 3/3
(U) AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH
K W BAUER MAY 87 AFIT/CI/NR-87-132D
UNCLASSIFIED F/G 12/3 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963-A

```
605   format(1x,'uncontrld coverage on sample mean of 1000 reps',f12.8)

      stop
      end

      subroutine gauss(ib,is,ip,a,kp,nvar)
c
c        THIS SUBROUTINE PERFORMS THE PIVOTS FOR VARIABLE
c             INTRODUCTION INTO REGRESSION MODELS
c                  FURNIVAL AND WILSON 1974
c
      real a(nvar,nvar,nvar)
      lb=ip+1
c
c        TOLERANCE CHECK ON PIVOTS
c

      if(a(ib,ip,ip).lt..01) then
       do 10  l=lb,kp
         a(is,ip,l)=a(ib,ip,l)
       do 10 m=l,kp
         a(is,l,m)=a(ib,l,m)
10      continue
       return
      else
       do 1 l=lb,kp
        a(is,ip,l)=a(ib,ip,l)/a(ib,ip,ip)
       do 1 m=l,kp
        a(is,l,m)=a(ib,l,m)-a(ib,ip,m)*a(is,ip,l)
1       continue
       return
      endif

      end


      subroutine keepit(numreg,nk,nx,models,knx,nvar)
c
c THIS SUBROUTINE FINDS THE MODEL OF A CANDIDATE REGRESSION
c
      integer nk(nvar),models(knx,nx)
      do 1 i=1,nx
       models(numreg,i)=nk(i)
1      continue
      return
      end
c
c  THE FOLLOWING FUNCTIONS ARE USED TO COMPUTE THE SELECTION
c                    CRITERION
c
```

```
      real function c3(k,iq,ip)
      c3=c1(k,iq,ip)*c2(k,iq,ip)
      return
      end

      real function c4(k,iq,ip)
      prod=1.
      do 10 i=1,ip
       top=float(k-iq-i)
       bot=float(k-iq-1)
       prod=prod*(top/bot)
10     continue
      c4=prod
      return
      end

      real function cfront(k,iq,ip)
      top=float(k-iq-1)
      bot=float(k-iq-ip)
      cfront=(top/bot)**ip
      return
      end

      real function c1(k,iq,ip)
      prod=1.
      do 10 i=1,ip
       itop=(k-iq-i)
       ibot=(k-iq-1)*k
       term=float(itop)/float(ibot)
       prod=prod*term
10     continue
      c1=prod
      return
      end
*
      real function c2(k,iq,ip)
      sum=0.
      p1=1.
      p2=1.
      do 10 j=0,ip
       ileft=jcomb(ip,j)
       if(j.ne.0) then
         p1=p1*(iq+2*(j-1))
         p2=p2*(k-iq-(2*j))
         rnext=p1/p2
       else
         rnext=1.
       endif
       term=float(ileft)*rnext
       sum=sum+term
10     continue
      c2=sum
```

```
      return
      end

      integer function jcomb(n,m)
      itop=nfact(n)
      ibot=nfact(n-m)*nfact(m)
      jcomb=itop/ibot
      return
      end
      integer function nfact(m)
      if(m.eq.0)then
       nfact=1
       return
      endif
      ip=m
      iloop=m-1
      do 10 i=iloop,2,-1
       ip=ip*i
10     continue
      nfact=ip
      return
      end
c
c THIS SUBROUTINE COMPUTES A F TABLE (TO THE POWER P)
c
      subroutine ftabl(ff,nx)
      common sig,kk,iqq,ip
      real root(1),last,ff(0:nx)
      external f
      eps=.001
      nsig=5
      nroot=1
      itmax=1000
      last=3.
      do 10 iqq=0,nx
       root(1)=last
101     call zreal2(f,eps,eps,eps,nsig,nroot,root,itmax,ier)
       if(ier.eq.33) then
        root(1)=last+1.
        ier=0
        write(6,102)
102      format(1x,'ignore last ier=33 warning --- reinitializing')
        go to 101
       endif
       last=root(1)
       fp=root(1)**ip
       ff(iqq)=fp
10     continue
      return
      end

      real function f(z)
```

```
      common sig,kk,iqq,ip
      n1=ip
      n2=kk-iqq-ip
      call mdfd(z,n1,n2,p,ier)
      f=sig-p
      return
      end

      subroutine cover(vcv,n1,nn1,models,knx,nx,iat,iin,ybar
     &,cbar,vecmuc,ny,vecmuy,numreps,ff,ih,icover,volred,vecybar
     &,iknow,covcv)
c
c  THIS SUBROUTINE DOES THE COVERAGE AND VOLUME REDUCTION CALC
c                FOR THE OPTIMAL CONTROL SUBSET

      parameter(nx2=7,ny2=2,n12=nx2+ny2,nn12=((n12*(n12+1))/2))

      real vcv(nn1),ybar(ny),cbar(nx),vecmuc(nx),vecmuy(ny)
     &,ff(0:nx),vecybar(ny),volred(2),covcv(nx,nx)
      integer models(knx,nx),ih(n1),icover(4)

      real scbar(nx2),svecmu(nx2),subv(nn12),subvf(n12,n12)
     &,b(n12),wkarea(2*n12),buff1(n12,n12),buff2(ny2,nx2)
     &,beta(ny2,nx2),cdev1(1,nx2),cdev2(nx2,1),expl(ny2,ny2)
     &,dev(ny2,1),ybhat(ny2),buff3(nx2,ny2)
     &,buff4(ny2,ny2),sydotc(ny2,ny2),hph(1,1),t1(1,nx2)
     &,ymd1(1,ny2),ymd2(ny2,1),t2(1,ny2),obs(1,1)
     &,buff5(ny2,ny2),buff6(ny2,ny2),ymd3(1,ny2),ymd4(ny2,1)
     &,obs2(1,1)
     &,symcovc((nx2*(nx2+1))/2),subcovc((nx2*(nx2+1))/2)
     &,fulcovc(nx2,nx2),gamma(ny2,nx2)
     &,ehat(ny2,ny2),buff9(ny2,ny2)
     &,cancorr(ny2,ny2),reigs(ny2),eigs(2*ny2),dummy(ny2,ny2)
     &,wk(ny2)
      integer ih2(nx2)
      complex ceigs(ny2)
      equivalence (eigs(1),ceigs(1))

c
c   INITIALIZE COVERAGE AND VOLUME REDUCTION VECTORS
c
      do 8 i=1,4
       icover(i)=0
8     continue
      do 9 i=1,2
       volred(i)=0.
9     continue

c
c   FIND THE SUBMATRIX FOR THE SELECTED MODEL
c
```

```
      do 10 i=1,n1
        if(i.le.nx) then
         ih(i)=0
         ih2(i)=0
        else
         ih(i)=1
        endif
10     continue

      ivar=0
      do 50 ii=nx,1,-1
        ivar=ivar+1
        if(models(iat,ii).eq.1) then
         ih(ivar)=1
         ih2(ivar)=1
        endif
50     continue

      m1=n1
      call rlsubm(vcv,m1,ih,subv,m2)

c
c     FIND THE SUBVECTOR (POPULATION AND SAMPLE) OF THE
c               CONTROL MEANS
c

      index=0
      do 100 ii=1,nx
        if(ih(ii).eq.1) then
         index=index+1
         scbar(index)=cbar(ii)
         svecmu(index)=vecmuc(ii)
        endif
100    continue

c
c      BUFFER THE COVARIANCE MATRIX OF SELECTED CONTROLS
c                AND RESPONSES
c
      call vcvtsf(subv,m2,subvf,n12)

      do 101 i=1,m2
      do 101 j=1,m2
         buff1(i,j)=subvf(i,j)
101    continue

c
c      INVERT THE COVARIANCE SUBMATRIX OF CONTROLS
c

      call linv3f(subvf,b,1,iin,n12,d1,d2,wkarea,ier)
      if(ier.ne.0)print *,"IDIED BELOW 101"
```

```
c
c      BUFFER THE CROSS-COVARIANCE SUBMATRICES OF SELECTED CONTR
c                     WITH RESPONSES
c
      do 102 i=iin+1,m2
      do 102 j=1,iin
        buff2(i-iin,j)=buff1(i,j)
        buff3(j,i-iin)=buff1(i,j)
102   continue

c
c      BUFFER THE COVARIANCE SUBMATRIX OF RESPONSES
c

      do 105 i=iin+1,m2
      do 105 j=iin+1,m2
       buff4(i-iin,j-iin)=buff1(i,j)
       buff6(i-iin,j-iin)=buff1(i,j)
105   continue

c
c      FIND THE BETA HAT MATRIX ( CONTROL COEFFICIENTS )
c      OR THE GAMMA HAT MATRIX
c

      if(iknow.eq.0) then

       call vmulff(buff2,subvf,ny2,iin,iin,ny2,n12,beta,ny2,ier)

      else

       call vmulff(buff2,subvf,ny2,iin,iin,ny2,n12,beta,ny2,ier)
       call vcvtfs(covcv,nx2,nx,symcovc)
       call rlsubm(symcovc,nx2,ih2,subcovc,iorder)
       call vcvtsf(subcovc,iorder,fulcovc,nx2)
       call linv3f(fulcovc,b,1,iin,nx2,d1,d2,wkarea,ier)
       call vmulff(buff2,fulcovc,ny2,iin,iin,ny2,nx2,gamma,ny2,ier)

      endif

c
c      FIND THE VECTOR OF CORRECTIONS TO CONTROL Y BAR
c

      do 103 i=1,iin
        cdev1(1,i)=scbar(i)-svecmu(i)
        cdev2(i,1)=cdev1(1,i)
103   continue

      if(iknow.eq.0) then
       call vmulff(beta,cdev2,ny2,iin,1,ny2,nx2,dev,ny2,ier)
      else
```

```
      call vmulff(gamma,cdev2,ny2,iin,1,ny2,nx2,dev,ny2,ier)
      endif
c
c     FIND THE CONTROLLED ESTIMATOR OF THE MEAN
c
      do 104 i=1,ny2
       ybhat(i)=ybar(i)-dev(i,1)
104   continue


c
c     FIND THE MATRIX OF EXPLAINED COVARIANCE DUE TO
c               CONTROL
c
      call vmulff(beta,buff3,ny2,iin,ny2,ny2,nx2,expl,ny2,ier)

c
c          FIND THE RESIDUAL COVARIANCE
c
      c1=(float(numreps-1)/float(numreps-iin-1))

      do 106 i=1,ny2
      do 106 j=1,ny2
       sydotc(i,j)=(buff4(i,j)-expl(i,j))*c1
       buff5(i,j) =sydotc(i,j)
106   continue
c
c         FIND THE ESTIMATOR SIGMA TILDE HAT
c
      if(iknow.eq.1) then

       const1=(float(numreps-2))/(float(numreps*(numreps-1)))
       const2=(float(iin+1))/(float(numreps*(numreps-1)))
       do 206 i=1,ny2
       do 206 j=1,ny2
        ehat(i,j)=(const1*sydotc(i,j))+(const2*buff4(i,j))
        buff9(i,j)=ehat(i,j)
206    continue

      endif

c
c        FIND THE INVERSE RESIDUAL COVARIANCE MATRIX
c
      if(iknow.eq.0) then
       call linv3f(sydotc,b,1,ny2,ny2,d1,d2,wkarea,ier)
      else
       call linv3f(ehat,b,1,ny2,ny2,d1,d2,wkarea,ier)
      endif

c
```

```
c          COMPUTE THE DEVIATIONS FROM THE STEADY-STATE
c                    RESPONSE VECTOR
c             (both cases: controlled/uncontrolled)
c
      do 107 i=1,ny2
        ymd1(1,i)=ybhat(i)-vecmuy(i)
        ymd2(i,1)=ymd1(1,i)
        ymd3(1,i)=ybar(i)-vecmuy(i)
        ymd4(i,1)=ymd3(1,i)
107   continue

c
c                  COMPUTE H'H
c        (Notation as per Venkatraman and Wilson 1986)
c
      if(iknow.eq.0) then
       call vmulff(cdev1,subvf,1,iin,iin,1,n12,t1,1,ier)
       call vmulff(t1,cdev2,1,iin,1,1,nx2,hph,1,ier)
      endif

      if(iknow.eq.0) then
       x=(1./float(numreps))+(1./float(numreps-1))*hph(1,1)
      else
       x=1.
      endif

c
c               COMPUTE THE RIGHT HAND SIDE
c                FOR THE CONFIDENCE REGION
c                  AS PER RAO (1967)
c
      c2=(float((numreps-iin-1)*ny2)/float(numreps-iin-ny2))
      f=exp((1./float(ny2))*alog(ff(iin)))
      rhs=x*c2*f

c
c               COMPUTE THE T**2 STATISTIC
c          FOR THE CASE WHERE  CONTROLS ARE USED
c                (steady state assumed)
c
      if(iknow.eq.0) then
       call vmulff(ymd1,sydotc,1,ny2,ny2,1,ny2,t2,1,ier)
       call vmulff(t2,ymd2,1,ny2,1,1,ny2,obs,1,ier)
      else
       call vmulff(ymd1,ehat,1,ny2,ny2,1,ny2,t2,1,ier)
       call vmulff(t2,ymd2,1,ny2,1,1,ny2,obs,1,ier)
      endif
c
c                  INDICATE COVERAGE
c          FOR THE CASE WHERE  CONTROLS ARE USED
c                (steady state assumed)
```

```
      if(obs(1,1).le.rhs) then
       icover(1)=1
      else
       icover(1)=0
      endif

c
c            COMPUTE THE VOLUME REDUCTION
c
      if(iknow.eq.0) then
       call linv3f(buff4,b,4,ny2,ny2,d1,d2,wkarea,ier)
       ucdet=d1*2**d2
       call linv3f(buff5,b,4,ny2,ny2,d1,d2,wkarea,ier)
       cdet=d1*2**d2
      else
       call linv3f(buff4,b,4,ny2,ny2,d1,d2,wkarea,ier)
       ucdet=d1*2**d2
       call linv3f(buff9,b,4,ny2,ny2,d1,d2,wkarea,ier)
       cdet=d1*2**d2
      endif

      term1=(cdet/ucdet)**(.5)*x**(float(ny2)/2.)
      c3=float((numreps-iin-1)*(numreps)*(numreps-ny2))
      c4=float((numreps-iin-ny2)*(numreps-1))
      term2=(c3/c4)**(float(ny2)/2.)
      f2=exp((1./float(ny2))*alog(ff(0)))
      term3=(f/f2)**(float(ny2)/2.)
      volred(1)=(1.-(term1*term2*term3))*100.

c
c              COMPUTE THE T**2 STATISTIC
c           FOR THE CASE WHERE no CONTROLS ARE USED
c
      call linv3f(buff6,b,1,ny2,ny2,d1,d2,wkarea,ier)
      call vmulff(ymd3,buff6,1,ny2,ny2,1,ny2,t2,1,ier)
      call vmulff(t2,ymd4,1,ny2,1,1,ny2,obs2,1,ier)

c              COMPUTE THE RIGHT HAND SIDE
c               FOR THE CONFIDENCE REGION
c
      c5=(float((numreps-1)*ny2)/float((numreps-ny2)*numreps))
      rhs2=exp((1./float(ny2))*alog(ff(0)))*c5

c
c                INDICATE COVERAGE
c           FOR THE CASE WHERE no CONTROLS ARE USED
c                (steady state assumed)
c
      if(obs2(1,1).le.rhs2) then
       icover(2)=1
      else
       icover(2)=0
```

```
      endif
c
c            THE REMAINING ANALYSIS DUPLICATES THE ABOVE SAVE THAT
c                THE GRAND MEAN OF 1000 RESPONSES IS USED
c
c                      RECOMPUTE DEVIATIONS
c
      do 108 i=1,ny2
        ymd1(1,i)=ybhat(i)-vecybar(i)
        ymd2(i,1)=ymd1(1,i)
        ymd3(1,i)=ybar(i)-vecybar(i)
        ymd4(i,1)=ymd3(1,i)
108   continue

c
c                COMPUTE THE T**2 STATISTIC
c            FOR THE CASE WHERE CONTROLS ARE USED
c                   (Grand mean used)
c
      if(iknow.eq.0) then
       call vmulff(ymd1,sydotc,1,ny2,ny2,1,ny2,t2,1,ier)
       call vmulff(t2,ymd2,1,ny2,1,1,ny2,obs,1,ier)
      else
       call vmulff(ymd1,ehat,1,ny2,ny2,1,ny2,t2,1,ier)
       call vmulff(t2,ymd2,1,ny2,1,1,ny2,obs,1,ier)
      endif

c
c                   INDICATE COVERAGE
c            FOR THE CASE WHERE  CONTROLS ARE USED
c
      if(obs(1,1).le.rhs) then
        icover(3)=1
      else
        icover(3)=0
      endif

c                COMPUTE THE T**2 STATISTIC
c            FOR THE CASE WHERE no CONTROLS ARE USED
c                   (Grand mean used)

      call vmulff(ymd3,buff6,1,ny2,ny2,1,ny2,t2,1,ier)
      call vmulff(t2,ymd4,1,ny2,1,1,ny2,obs2,1,ier)

c
c                   INDICATE COVERAGE
c            FOR THE CASE WHERE no CONTROLS ARE USED
c
      if(obs2(1,1).le.rhs2) then
        icover(4)=1
      else
        icover(4)=0
```

```
      endif
c
c           THIS SECTION COMPUTES THE CANONICAL CORRELATIONS
c           FOR THE SUBSET MODELS AND THE FEASIBILTY BOUND
c           FOR USING THE KNOW COVARIANCE MATRIX OF CONTROLS
c
      if(iknow.eq.1) then
       call vmulff(buff6,expl,ny2,ny2,ny2,ny2,ny2,cancorr,ny2,ny2,ier)
       call eigrf(cancorr,ny2,ny2,0,eigs,dummy,ny2,wk,ier)

       icount=0
       do 300 i=1,ny2
       do 300 j=1,2
        icount=icount+1
        if(j.eq.1) reigs(i)=sqrt(eigs(icount))
300     continue

       ctop=float((numreps+iin-1)*(numreps-iin-2))/
     &      float((numreps-1)*(numreps-2))
       cbot=ctop*(float(numreps-2)/float(numreps+iin-1))
       bound=sqrt((ctop-1.)/(cbot-1.))
       print *,"Canonical correlations ",reigs," Bound ",bound
       print *,eigs
      endif

      return
      end
c
c          THIS SUBROUTINE RETURNS THE GENERALIZED VARIANCE
c                 OF SIGMA TILDE HAT
c
      subroutine covknow(rss,ny,full,nvar,target,dum,numreps,mv,det)
      real rss(ny,ny),full(nvar,nvar),target(ny,ny),dum(ny)

      c1=(float(numreps-2)/float(numreps*(numreps-mv-1)))
      c2=(float(mv+1)/float(numreps*(numreps-1)))
      nx=nvar-ny

      do 10 i=1,ny
      do 10 j=1,ny
       target(i,j)=(c1*rss(i,j))+(c2*full(nx+i,nx+j))
10     continue

      call linv3f(target,dum,4,ny,ny,d1,d2,wkarea,ier)
      det=d1*2**d2
      return
      end
```

VITA

VITA

Kenneth W. Bauer, Jr. was born on 15 March 1954 in Xenia, Ohio. He recieved a B.S. in mathematics from Miami University, Ohio in August 1976. He recieved a M.E.A. (Engineering Administration) from the University of Utah in June 1980. He recieved a M.S. in Operations Research from the Air Force Institute of Technology in December 1981.

In August 1984 he entered the Ph.D. program in Industrial Engineering at Purdue University.

He is currently a Captain in the United States Air Force serving as an Assistant Professor at the Air Force Institute of Technology.

END
DATE
FILMED
JAN
1988